上海出版资金项目

Shanghai Publishing Funds

主　编　王思斌　曾华源

执行主编　沈　黎

简春安

邹平仪

著

社会工作研究方法（下）

华人社会工作丛书

（第一期）

华东理工大学出版社

VISIT…

LANZAROTE
Caliente.com

主　　编　王思斌　曾华源
执行主编　沈　黎

简春安
邹平仪
著

社会工作研究方法（下）

华人社会工作丛书（第一期）

華東理工大學出版社
EAST CHINA UNIVERSITY OF SCIENCE AND TECHNOLOGY PRESS
·上海·

图书在版编目(CIP)数据

社会工作研究方法(下)/简春安,邹平仪著.—上海:华东理工大学出版社,2018.6
(华人社会工作丛书.第一期)
ISBN 978-7-5628-5327-5

Ⅰ.①社… Ⅱ.①简… ②邹… Ⅲ.①社会工作-研究方法 Ⅳ.①C916-3

中国版本图书馆CIP数据核字(2017)第319950号

内容提要

本书内容分为研究计划、研究方法、资料分析三部分。本书对理念或概念的解析都很明确独到,且于每个概念后加上了生活化的实例介绍,使理论与实务合二为一。本书体系完整,叙述深入浅出、说理清晰、例证贴切,易读、易懂、易学。本书不仅可以作为社会工作及其相关科系的教材,且对参加各种考试及有兴趣研究者也很适用。

策划编辑/刘　军
责任编辑/刘　军　孟媛利
装帧设计/徐　蓉
出版发行/华东理工大学出版社有限公司
　　地址:上海市梅陇路130号,200237
　　电话:021-64250306
　　网址:www.ecustpress.cn
　　邮箱:zongbianban@ecustpress.cn
印　刷/山东鸿杰印务集团有限公司
开　本/710 mm×1 000 mm　1/16
印　张/14.5
字　数/226千字
版　次/2018年6月第1版
印　次/2018年6月第1次
定　价/78.00元

版权所有　侵权必究

学 术 委 员 会

（按姓氏笔画排序）

中国大陆

马凤芝教授（北京大学）
王思斌教授（北京大学）
文军教授（华东师范大学）
史柏年教授（中国青年政治学院）
关信平教授（南开大学）
李迎生教授（中国人民大学）
何雪松教授（华东理工大学）
顾东辉教授（复旦大学）
徐永祥教授（华东理工大学）
彭华民教授（南京大学）
蔡禾教授（中山大学）

中国香港

马丽庄教授（香港中文大学）
庄明莲教授（香港城市大学）
陈丽云教授（香港大学）
梁祖彬教授（香港大学）

中国台湾

古允文教授（台湾大学）
莫藜藜教授（东吴大学）
黄源协教授（暨南国际大学）
曾华源教授（东海大学）

海外地区

齐铱教授（美国南加州大学）
陈伍强教授（新加坡新跃社科大学）
张锦芳教授（美国休斯敦大学）
殷妙仲教授（加拿大不列颠哥伦比亚大学）
梅陈玉婵教授（美国哥伦比亚大学）
梁毓熙教授（美国休斯敦大学）
韩文瑞教授（美国纽约大学）
曾家达教授（加拿大多伦多大学）

编 辑 委 员 会

主　　编：王思斌　曾华源

执行主编：沈　黎

丛 书 总 序

北京大学社会学系　王思斌

随着我国社会转型和现代化进程的加快，民生问题、社会治理问题更加突出地摆在政府和社会面前。社会工作作为帮助弱势群体解决其基本生活困难、走出困境，进而促进社会公平正义和社会进步的专业学科，其快速和稳步发展显得十分重要。《中共中央关于构建社会主义和谐社会若干重大问题的决定》作出“建设宏大的社会工作人才队伍”的战略部署，中央十九部委也制定了《社会工作专业人才队伍建设中长期规划（2011—2020 年）》。此外，在政府、学术界和实务界的共同努力下，我国的社会工作事业得到了较快发展。但是，与解决上述问题的迫切性和社会工作学科发展的要求相比，我国大陆的社会工作发展还存在很多不尽如人意之处，其中一个重要原因就是社会工作专业化与本土化的结合还不到位。

社会工作产生于西方发达国家，其基本理念、工作方法带有浓厚的西方文化色彩，带有它产生于其中的社会结构的印记。在改革开放和走向现代化的进程中，我国要发展社会工作事业，需要借鉴西方社会工作的优秀成果，走专业化之路。但是，我们必须同时清醒地认识到，在此过程中照搬西方社会工作的既有成果有时并不恰当，因为我国的经济体制、政治体制、文化体制、社会体制与西方有很大差异。在这种情况下，社会工作本土化就显得尤为重要。关于社会工作本土化，学界有许多不同的理解。但是如果说来自西方的社会工作要适应我国的基本国情和解决问题的具体环境，则是没有那么多争论的。来自西方的社会工作理论、工作方法必须在中国能落地，能与需要解决的现实问题相契合，这是问题的关键。正是在这种意义上，笔者曾经讨论过社会工作本土化问题，也提出了社会工作本土化的几种路径：第一，研究西方社会工作与中国国情的

不相容之处，删去不适宜的部分，再应用于中国实践，并接受实践检验；第二，对从事实际社会工作的人员进行专业培训，促进国际社会工作经验与本土社会工作的比较和碰撞；第三，分析两种社会工作之异同，找出它们的相似之处，强化它们之间的亲和性；第四，在此基础上从事社会工作实践，扩大两者之间的相容性，以实现优势互补。在上述可能的实践中，笔者还提出过一种看法：本土社会工作学者在其中能发挥重要作用。因为，一方面，他们熟悉产生于西方的社会工作理论和方法；另一方面，他们长期生活于非西方社会之中，对何种社会工作理论、知识和方法有助于本地问题的解决有比较清楚的认识，从而进行选择性“过滤”。

回到当下我国社会工作的现实处境。虽然大陆的社会工作学科重建已有30年，但是在学术研究、经验总结、理论反思方面还相对滞后。相比之下，我国港台地区的社会工作专业发展较早，实践比较充分。在来自西方的理论与本地实践的磨合互动中，一些学者形成了自己的理论观点、经验模式、知识积累，这对同具华人社会文化的大陆地区来说具有重要的参考价值，也是华东理工大学出版社出版这套“华人社会工作丛书”的重要初衷。本丛书选择工作生活于世界各地（主要是欧美和我国港台地区）的资深华人社会工作学者的有分量的著述加以印行，在某些方面和一定程度上能缩短大陆地区社会工作本土化的进程，也有利于向读者提供一个参照——这些优秀的学者是怎样在专业化与本土化之间进行思考的。当然，这些学者工作、研究和进行专业实践的背景与大陆地区也不相同，政治制度、社会体制、社会福利体系和社会工作制度方面的差异也比较明显，因此对他们的理论、观点我们也不能照抄照搬。如果大陆读者能在这些华人学者的理论、经验与自己面对的现实之间进行比较、思考和反思，有了新见解，那就真正从深层次上达到出版本丛书的目的了。说不定这会对华人社会工作乃至世界社会工作的发展有所贡献。但愿如此。

序

用浅白的文字写一般人轻易就可以看懂的研究法教科书，是我多年来的心愿，1998 年出版的时候如此，2004 年修订时如此，2013 年退休离开教职和 2016 年进行最后一次修订，亦复如是，算是对社会工作学术界的告别。

把研究法的书写得简单易懂，虽然得到很多读者的欢迎，但写作过程也因此艰辛。因为把理论与抽象的概念写成白话，甚至用生活上的例子来说明时，马上就失去了文字表达的安全空间；而研究法里一些至今仍模棱两可的定义、见仁见智的说法，就算能说得浅白，也难表达得淋漓尽致。用这个方式来写书需要不小的勇气。所幸，这个勇气仍然存在，年轻刚教书时如此，现在退休了，仍然如此。

谢谢邹平仪博士对本书第十三章的贡献，把社会工作实务同仁平日所惧怕的统计简要地做了介绍。也谢谢高雄医学大学的陈政智教授，在文献综述中对如何在网络上收集资料做了很实用的说明，但愿对社工伙伴们在准备考试或自己做研究时有实质性的帮助。更重要的是，他对本书教学 PPT 的制作耗费了不少心力，若因此让更多学校教师更愿意采用本书，这将是陈政智老师的贡献。

1998 年本书初版时，有助理方婷像天使般的协助。2013 年退休了，我没有助理，试着孤军奋斗，因此所造成的瑕疵或编排上的错误，在所难免，敬请读者们谅解。谢谢东海研究所的学生在上课时给我的意见，谢谢东海社工系同仁在书目格式单元中的建议，使本书增色不少。

再一次把本书献给在实务界不断付出的社工朋友们,若能因此对各位的进修、升学或社工师考试有所助益,那将是我最大的喜乐。也期待社工学生们,把研究法读好,研究时不离实务,实务工作时不忘研究,让社工更进步,让一代胜过一代。

简春安

目　　录

第八章　问 卷 设 计

第一节　前言 …… 003
第二节　测量的定义 …… 005
第三节　量度的性质 …… 011
第四节　量表设计的方式 …… 012
第五节　问卷的形式 …… 017
第六节　问卷的信度测量 …… 025
第七节　问卷的效度测量 …… 028
第八节　总加量表的鉴别力(*DP* 值)分析 …… 032
第九节　结论 …… 034

第九章　实 验 研 究

第一节　前言 …… 039
第二节　实验研究的类型 …… 040
第三节　“实验”的内涵 …… 041
第四节　实验设计(研究设计) …… 045
第五节　一般常见的实验设计 …… 053
第六节　单案研究设计 …… 065

第十章　其他常见的研究法

第一节　前言 …… 073
第二节　文化比较研究法 …… 073
第三节　历史研究法 …… 074
第四节　再次分析法 …… 080
第五节　内容分析法 …… 086
第六节　非干扰测量 …… 089
第七节　结论 …… 091

第十一章　事后回溯研究与方案评估

第一节　事后回溯研究法 …… 095
第二节　方案评估 …… 100
第三节　德怀术(Delphi panel)研究法 …… 114
第四节　行动研究 …… 117

第十二章　抽　　样

第一节　前言 …… 123
第二节　抽样的概念与术语 …… 123
第三节　如何进行随机抽样 …… 126
第四节　抽样 …… 129
第五节　结论 …… 135

第十三章　资料分析方法

第一节　前言 …… 139
第二节　统计方法在社会研究资料分析中的功能、限制与选判原则 …… 140

第三节　资料分析的统计方法 …… 142
第四节　实务研究与资料分析的其他议题讨论 …… 168

第十四章　论文写作

第一节　前言 …… 183
第二节　研究论文的次序与格式 …… 183
第三节　论文第三章——“方法论”的写作 …… 186
第四节　论文第四章——“分析与发现”的写作 …… 191
第五节　论文第五章——“结论”的写作 …… 197
第六节　参考文献与附录 …… 199
第七节　其他 …… 200

参考文献 …… 202

第八章 问卷设计

第一节　前言
第二节　测量的定义
第三节　量度的性质
第四节　量表设计的方式
第五节　问卷的形式
第六节　问卷的信度测量
第七节　问卷的效度测量
第八节　总加量表的鉴别力(*DP* 值)分析
第九节　结论

摘　要

问卷不是研究者想到哪里就可以问到哪里。问卷设计是研究过程中极其重要的部分，研究的概念如何落实，资料能否有效被收集，如何借着研究中的假设及其所牵连的变量，把理念转变为一些相当具体的题目，以测量所要调查的变量……这些都需要在问卷设计上下功夫。

量表是借着一些题目来测量，不同的性质在量表中有不同的功能，也因此在资料分析时必须采用不同的统计方法，所达到分析的功能亦会有程度上的差异。一般而言，量表可分为类别、顺序、等距、等比四种。量表的设计方面，则以总加量表、等距量表、累积量表最为普遍，其制作过程将于第四节详述。

好的问卷首先必须注意到每个题目的用词与布局，面对不同的受访者，理想的问卷当然必须面面俱到。从理论上来讲，好的问卷最重要的是问卷的信度与效度，若是在信度、效度上能够清楚交代，调查几乎已成功了一半。一般而言，信度可分为再测信度、折半信度、分数信度、复本信度四种；效度可分为内容效度、并行效度、效标效度、建构效度等几种，如何进行问卷的信度、效度分析，读者不可不慎！

第一节　前　　言

社会调查过程中，很多时候都要用问卷去测量，问卷的设计是否准确、合宜，将大大地影响调查的结果。问卷绝对不是想问什么就问什么，也不是研究者想问几个问题就问几个问题；问卷乃是一项严谨规划的过程，问卷的来源、结构，以及组合都应有相当的理论基础，也需要仔细的设计，否则问卷在研究中便失去了意义（Feick，1989）。问卷是借着研究中的假设，及其所牵连的变量；或对抽象的理论不断地概念化与操作化的过程，并把理念转变为一些相当具体的题目，用之于测量所要调查的变量，兹分述如下。

一、概念（concept）的意义

概念指对某件事物的定义。例如，“成就感”是个概念，“满足感”也是个概念，我们可以从字典中为这两个概念找到定义。较为专业的名词，就在较专业的字典或百科全书中寻找，把这个概念定义清楚。问题是，不同的字典、不同的行业，或不同的个体，对同一个概念可能会有不同的定义。有的定义简洁了断，有的则是周详清楚；有的偏向生理科学，有的则偏向心理……这些差异导致一些概念存在这样的问题：表面大家以为彼此看法一致，等到详细探讨时，才发现其实看法各有差异。到底彼此之间是相同还是差异？若是相同，是如何的类似？若是差异又是如何的不同？做研究时，应该仔细分析，马虎不得。定义其相同或差异最简单的方法，就是设法对其概念予以操作化。

二、操作化（operationalization）

把抽象的定义予以实体化，使之可以用测量的方式来衡量此概念时，即谓操作化。在对一个概念进行操作化时，常会因该概念的性质而有不同的操作化过程。有时会用实际定义（real definitions）的方式来进行操作化，字典怎么说，定义就怎么做，该定义也是该概念的操作化。例如，“无助感”（helplessness）这

个概念,若字典的定义是"觉得对事情无能为力,缺少改变的动机、企图与技巧",那这个定义就是该概念的实际定义。

名义定义(nominal definitions)是另外一个操作化的方法,由研究者对该概念给予一些解释与定义。例如,同样是"无助感",名义定义可能由研究者对"无助感"给予的三种说明而形成:(1)无望感(hopelessness),对某事不抱希望;(2)无意义感(meaninglessness),觉得什么事都没有意义;(3)无力感(powerlessness),凡事都提不起劲,不认为自己可以改变现实,认为渺小的自我对外界不可能产生任何影响。此时无助感已被更清楚地界定为无望感、无意义感、无力感等三个方面的内容,这就是名义定义。

一般讲的操作化是指操作性定义(operationalizing definition)。再以"无助感"为例,它的操作性定义是:

> "无助感指'生活满意度调查表'所测得的分数;'行为改变动机量表'所测得的分数;加上'企图心指数量表'的分数的总和。当分数愈高时,其无助感的程度就愈低。"

有了这个操作性的定义,"无助感"这个概念不仅已经定义清楚,而且可以用分数来呈现,这就是操作化及测量。换句话说,操作性定义就是把抽象的概念运作到可以测量,甚至可以用分数来计算的地步。

变量本来都只是一个概念,为了研究及分析的方便,研究者必须把此概念以分数的形式来呈现,此过程即为变量的测量。如果变量无法测量,研究时的实验组和控制组之间到底孰优孰劣便无法比较,有关的假设当然也就无法检定。当然研究者也必须牢记,测量并非研究的目的,也不是整个研究的目标,测量只是达到目的的手段,研究时千万不要倒果为因;当变量无法测量时,或是测量的信度与效度遇到极严重的瓶颈时,就应在研究的过程中仔细交代,使读者知道其研究的限制,而不要为了报告如期完成,或是反正没人会注意到测量的制作过程,而有草率或大意的行为。如果这样,纵使研究有了结果,却是与事实不符的发现,只会徒然使专业产生错误而已。

第二节 测量的定义

一、测量的定义

“测量”从字面上的定义是 Measurement is the assignment of numbers to objects or events according to certain rules(Steven, 1951)。换句话说,“测量是根据某些准则,把一些对象或事件赋予数字的过程”。从这个定义来分析,测量有三种含义:第一,测量是“赋予数字”(assigning number)的行为过程,把所观察到的东西赋予数字,把所研究的对象或概念予以量化;第二,测量必须有某些准则为依据(according to certain rules),凡是符合某种现象的,所被赋予的数字就必须相同,一切有准则为依据,不能光由研究者的好恶来决定;第三,测量的对象是所研究的对象或事件(objects or events)。

严格地说,测量的主要对象是研究者的自变量与因变量,这些变量被转化为一些题目(items),而有某些题目的总和就成为该变量的分数。由以上三个测量的项目去分析,我们可以归纳测量的过程其实就是:(1) 下定义(define a phenomenon),把所要观察的对象与赋予数字的方法定义出一些准则;(2) 做观测(observe),对对象做观测;(3) 配数字(select the number),使所观察的现象都能以数字来表示。

二、由理论衍化为问卷题目的流程

从研究的角度来看,问卷的来源牵涉到研究的目标,关联到研究根源的理论,更与研究的假设紧紧相扣。因此,问卷的形成绝对不是临时起意,也不是偶然的动作与行为。在林南(Lin, 1976)的《研究调查基础》(*Foundation of Social Research*)一书中,他仔细地陈述理论如何演化为一些与该理论相关的概念,概念与概念之间又如何形成一些与该理论相关的命题(propositions),这些概念如何转化为变量(variables),而变量与变量之间又如何形成假设(hypotheses),逻辑的演绎与归纳又如何在抽象的理论与具体的实务中穿梭(由抽象的理论变成具体的实务是演绎,由具体的实务变成抽象的理论则是归纳),这些理念与本章所要陈述的

“问卷的建构”都有密切的关系。具体而言,问卷的题目始于抽象层次较高的理论,理论内一定包含着不少相关的概念,每个概念又可以演绎为与研究较为有直接关系的变量,变量经过操作化以后,则变成可供受访者回答的题目,每个题目都有不同层次的回答,其答案则可换算为分数,分数又可供研究者利用统计的方法予以分析,研究的答案于是一步步地被寻找出来,新的发现也因此不断地被提出。

研究者必须注意的是,这种由理论而概念,由概念而变量,由变量而转变为一些问题(items)的过程,并不是一成不变的。因为某个研究者心目中的“变量”,可能是另一个研究者心目中的“概念”;也因为研究的目的或角度不同,变量与概念的划分也就不那么容易。例如,“满意程度”是概念?或是变量?可能因研究目标的不同而有差异。某甲研究“适应”,可能把“工作满意度”当作他的研究的一个“变量”。但是某乙专门研究“满意度”,因为“满意度”对他太重要也太复杂了,所以对他而言却是一个“概念”,而不是“变量”而已。总而言之,概念与变量之间不是单纯的“一步之隔”,端看研究者的研究方针或研究目的而定。

这种由抽象的理论演绎为较具体的变量或题目的过程,下面两例可供参考。

在吴秀照的硕士论文中,主要探讨“工作态度”与“满意度”之间的关系。而她对“工作态度”有一组较为具体的定义,她认为工作态度可包括组织融入、工作投入、工作士气及对社会工作的看法等四个主要的项目,如图 8-1 所示:

- 工作态度
  - 组织融入
    - 对组织的向心力
    - 对组织未来发展的关心
    - 为组织努力的程度
    - 与组织目标符合的程度
  - 工作投入
    - 留职倾向
    - 工作时间
    - 工作效率
    - 团队的参与性
  - 工作士气
    - 对主管的顺从
    - 对工作角色的自我评价
    - 对同事的信心
    - 在工作上的成就感
  - 对社会工作的看法
    - 职业喜好
    - 社会工作未来的发展性
    - 专案理念的认同
    - 专业伦理的肯定

图 8-1 “工作态度”的操作化

在以上的概念流程中,“工作态度”是研究者的重要“自变量”,但是从变量的层次来看的话,应该说是研究的“概念”,因为她对“工作态度”还有很多较为具体的演绎步骤。“组织融入”“工作投入”“工作士气”与“对社会工作的看法”可以说是“变量”,但仍然可以说是“概念”;说它是变量,因为它的上面是“工作态度”这个概念,说它是个概念,是因为它的下面还有四个变量,如对组织的向心力、对主管的顺从、职业喜好、专业理念的认同、专业伦理的肯定等。可见概念与变量之间尚无法截然划分,端看研究者如何去处置而已。但是不管研究者如何处置,概念的抽象层次一定是较变量为高;概念愈操作化,势必愈加具体,这乃是基本原则。

再以卓春英的论文中所探讨的“非医疗因素”为例,她的硕士论文以医疗过程中非医疗因素对病患的影响为主题。“非医疗因素”是其主要的自变量,我们可以把它当作一个概念来看,在医院里,医生对病人所做的是医疗行为,社会工作者在医院里也有其非医疗性的治疗功能。病人的生理疾病需要医生来医治,但是一些非医疗性的问题,则需要社会工作者来处理,否则医疗效果便会有折扣。为了证明这个特质,卓春英把非医疗因素操作化为一些变量,各个变量又可细分为一些较细、较具体的小变量,然后再把这些小变量设计为一些问卷题目,如图 8－2 所示:

- 家庭问题
  - 亲子关系
  - 夫妻问题
  - 家庭互动
- 职业问题
  - 失业
  - 就业困难
  - 劳资纠纷
- 经济问题
  - 医疗费用
  - 生活费用
  - 家庭开支
- 心理问题
  - 焦虑
  - 害怕
  - 担忧
- 社会问题
  - 人际关系
  - 顺从规范
  - 住院适应

图 8－2　“非医疗因素”的操作化

三、概念衍化过程中的几个考虑因素

问卷的来源根源于理论的衍化,由理论而概念,由概念而变量,由变量而小变量[小变量可称为某些“标准”(criteria)或“定义”(definition)],再由小变量衍化为一些可由受访者回答的问题。这种由概念逐步衍化为题目的过程虽然烦琐,但是若能由此原则来设计,却可保证问卷的理论效度或内容效度。刚开始学习设计问卷的同学若依此原则来设计问卷,就算烦琐,也不致手忙脚乱。但是这种设计的过程,仍然要注意一些事项。

(一)要注意衍化后的周延性(exhaustiveness)

由概念衍化成一些变量,好坏与否端看这些变量是不是能够表达概念。以“台湾地区”这个概念为例,变量中就必须包括北、中、南、东等四区,若还觉得不足,或是个案性质特殊(如要探讨城乡差距时),那离岛可能还要加进去。如此北、中、南、东及其他(离岛)五个变量加起来,在表达“地区”这个概念时,就大大增加了其周延性。吴秀照论文的“工作态度”这个概念,包含着组织融入、工作投入、工作士气及对社会工作的看法,大体也包含了工作态度所应包含的变量,也可以说它的周延性还不错。卓春英的“非医疗因素”中,病人的社会问题包括了人际关系、顺从规范,以及住院适应等三类,从一般的社会问题的定义来看,好像不是很周全,但是从住院病人的辅导经验来看,住院病人的社会问题大体上也是这三个问题,所以这三个定义的周延性仍然不错。

(二)互斥性也不可忽略(exclusiveness)

“台湾地区”概念中的北、中、南、东及其他(离岛)五个变量,只要把各区的地理范围规范清楚,各区有各区的范围,且不会重叠时,就表示它们之间的互斥性很好。若未能定义清楚,在分类或资料整理时就容易造成偏差,如嘉义应该属于中或南?宜兰属于北还是东?这样一来,资料就无法被客观分析。简春安在《外遇的分析与处置》一书中把外遇的类别分为传统型、拈花惹草型、保护型、情境型、感性型等几种类型,虽然每一型都有其定义,不致与其他类型重叠,但在实际的个案中,却有很多外遇的个案都由一种以上的因素所造成,可能是既感性、又保护;有的则是传统因素与情境因素的综合,如此一来,我们就很难把案主归到其中某一类型中,这种分类的互斥性就不是很完美,需要再予改进。

(三)同质性问题最好讲究

“台湾地区”的概念中,北、中、南、东的性质都相同,“离岛或其他”在语意上可能就与前四区不大搭调,虽然还不至于太离谱。心理问题中的害怕、焦虑与担忧,在英文的语意上同质性颇高,但是万一再加上“常常莫名其妙地觉得不舒服”,最后一个与前几个比起来可能就显得格格不入。根据某个概念设计变量时,这些变量最好属于同级的语意,使它的抽象程度大致相等。如此,在问卷的设计时,就可以使问卷的品质得以保证。

四、问卷题目设计应注意的事项

设计问卷时,若能事先掌握原则,不仅能轻松有效地完成设计,更能使研究者在分析时条理分明,处理统计时也较能得心应手、剑及履及。Babbie(1992)认为一些事项应该注意。

(一)问题与命题(questions and statements)

正规的研究调查所用的问卷,其背后一定有很多的研究命题,绝对不是随便或碰巧把一些问题集中在一起而已。研究者在规划问卷时,一定要根据研究命题来设计,问卷中的问题须紧紧与命题联系在一起;须根据命题来设计题目,并保证每个题目都有命题为依据。

(二)开放式(open-ended form)与封闭式(closed-ended form)的问题

开放式的问题可以让受访者有较大的空间回复,规划时也较简单,反正把问题列出来即可;但是资料的整理可能会累死人,因为开放式的问题常常无法对答案进行规格化的分析,在统计上更是寸步难行。封闭式的问题则把受访者所能回答的结果都限制在研究者所要求的范围里,反正一切答案都在所规划的范围内,大不了在四五个固定答案后面多一个“其他”,再画个横线,让受访者把心中所想的答案填进去。其好处是资料回收后很容易统计分析;缺点是把受访者的回答规划得太简单。一般而言,量化研究大都采用封闭式的问题,质性研究则必须以开放式的问题来进行,否则得不到研究者所要的资料。

(三)务必清楚了解问卷的题目与项目

当访问正在进行时,一些访问者极为清楚的事项,却可能会使受访者坠入五里雾中,搞不懂如何回答你的问题。例如,问他们“沟通的频率”时,他可能以为你是在问他们沟通的品质,而以他们“两人没话可讲”来回答;事实却是他们

平常沟通还算良好,只是昨晚吵了一次架而已("没话可讲"是指从昨天以来关系不好,却不是指夫妻谈话的量不多)。访问者听到这种回答时,最好再澄清一次,把受访者的答案填到适当的问题中,然后再问清楚他们沟通的频率到底如何,但是要做到这种地步,访问前非充分掌握问卷的项目与题目不可。

(四)避免一个题目中载有双重题意

封闭式的问卷中,每一题都只有一个答案,但若是题意中同时表达了两种信息,会使受访者不知从何作答。例如"孩子不打不成器,你赞成学校的体罚教育吗",受访者回答时不知是要赞同你的"孩子不打不成器"这个观念或是对"学校的体罚教育"的态度。

(五)有关题目的背景信息应该完全提供,使受访者方便作答

询问受访者对"虐待儿童"的态度时,必须先把什么叫作"虐待"交代清楚。要问是否赞成离婚时,应该把"何种状况"下的离婚或离婚的类别等事先说明,毕竟婚姻的状况不同,是否应该离婚的意见也会有差异。

(六)问卷题目之间应相互连贯;每个题目都应与主题有关

问卷的形成本来就应该根据研究的主题而来,绝对不是随兴写出一堆题目即可,好的问卷应该都是依据主题一气呵成,各个变量都能被规划成问卷的题目。同样地,相同的单元中,其题目应该是彼此相关的,若是相同单元中的题目不相关,则不仅其信度不高,问卷设计也一定会犯错。

(七)题目应简洁扼要

有效得到受访者答案的方法是把问题设计得尽量简洁扼要,千万不要冗长繁复。若问题无法简要表达清楚,常常是研究者在概念上也不很清楚所致。题目简洁扼要可使受访者在填答时觉得有成就感,不容易产生倦怠而中途放弃。

(八)避免负向的题目

在十几年前的研究法教材中,常常提醒研究者在设计题目时应该有一半是正向题,一半是负向题,而且建议正向题与负向题应该交叉叙述,当时认为如此安排可以使受访者"清醒"作答,其实效果却适得其反,徒然使受访者混淆不清而已。例如,问受访者"政府应该禁止人格不成熟的家长抚养在学校已经行为偏差的小孩",就算这一题没有造成太大的困扰,下一题再问"管教子女是家长之责,就是对管教子女成效不佳的父母,政府也无权干涉",如果之后的题目都是一正向一负向地交叉询问,受访者若要完成一连三十题甚至更多的题目,就

是不晕头转向，也一定是元气大伤。

（九）避免带有偏见或情绪化的题目

若要研究受访者对女权运动的态度，一般的题目如“天赋人权，女性也应该拥有跟男人一样的工作权”，回答这个问题时，受访者或可以平静的心情来作答。但是“天赋人权，女人若要与男人争工作机会时，你会赞成吗？”此题中，“与男人争工作机会”可能煽动性就强一点，男性回答这一题时，非常赞同的可能性必然降低。偏见、情绪化、太过主观、太极端的字眼等都会使受访者在作答时失去冷静的态度。

第三节　量度的性质

量表是借着一些题目（items）来测量，题目则是由变量转换而来。由变量转换为题目的过程中，会因变量的性质而造成不同量度（scale）性质的题目。不同的性质在量表中有不同的功能，也因此在资料分析时必须采用不同的统计方法，所达到的分析的功能亦会有程度上的差异，所以研究者在设计问卷时，针对题目的用语、所牵涉的变量的性质都应细察。有关量度的性质可参考表 8-1：

表 8-1　量度的性质

尺度类别	功　能	说明及例子
类别尺度（nominal）	分类	如生理性别可分为男、女两种。其他如宗教、婚姻状况、职业类别等都是类别尺度
顺序尺度（ordinal）	分类、排序	如受教育程度可以分为不识字、小学、初中、高中、大学、研究生及以上等
等距尺度（interval）	分类、排序、设定标准距离	如年代，公元前 1000 年、公元前 500 年、公元元年、公元 500 年、公元 1000 年，以及公元 1500 年、公元 2000 年等。其他如华氏温度等
等比尺度（ratio）	分类、排序、设定标准距离、含绝对零	如收入、年龄、重量、距离等。这种量度的性质与等距尺度之间的差别只在于有无绝对的 0。收入没有就没有，不像年代，公元〇年是概念而已

以上这四种量表性质不同，所能表现的功能境界也相异，而且在资料分析

时也会造成很大的差异。类别尺度只能分类,顺序尺度除了分类以外还可以排序,等距尺度则还能设定标准的距离(次序与次序之间可以看出量或质的不同),等比尺度则可以有绝对 0 的概念。在资料分析时,面对自变量与因变量之间复杂的关系,若只要看出其不同就可以时,我们当然可以用类别尺度;但若要排出其次序时,顺序尺度就绝对必要,反正研究者所求也只是类别与类别之间的关系状况而已;若要算出其差异的量,找出变量与变量之间一对一的线性关系时,那就非用等距或等比尺度不可了。

设计问卷时,研究者当然要了解到“尺度”不同时,准确度也当然不同。因此,在设计问卷题目的用词中,就必须选择适当的尺度量表。以教育程度为例,若我们所求是有没有受过学校教育,则“有无受过学校教育”的类别式问卷就已足够;若要看出教育水准的高低,则“不识字、小学、初中、高中、大学、研究生及以上”的顺序尺度的方式是必要的;若要精确地算出受教育年数的详细差异,则学校教育到底是几年的等距或等比尺度方式的问卷绝对必要,问卷则要问学校教育几年:从无,1,2,3,4,…,20,21,22 年以上等都需要调查出来。从这个角度来看,有经验的问卷设计者在询问教育程度时,他宁愿以受访者的学校教育年数到底是几年的等距或等比尺度的方式来问,而不会用类别的方式来处理,主要原因是,等距或等比尺度的方式可以转换为类别或顺序尺度,而类别或顺序尺度却不能转化为等距和等比尺度。

第四节　量表设计的方式

量表的设计有不同的方式,一般而言,下列三种最为普遍:总加量表;等距量表;累积量表。兹分别说明其制作的过程如下。

一、Likert 总加量表

李克特量表(Likert scale)是总加量表中最普遍的一种形式,其在社会调查中使用最为广泛,它的制作过程有下列几个步骤。

(一)总加量表是测量受访者“态度”时最容易使用的方法

基本上,总加量表是由一组专门调查受访者的态度、意见或看法的题目所

组成(平日的调查,有的调查行为的频率、方式;有的调查则专精于概念的类别,但是总加量表则是以态度、意见为主)。

（二）在这组量表中,每一个题目的分量都是等值的

例如,在一个题量为 20 的婚姻暴力研究问卷中,询问“夫妻曾否口角?”的题目与“夫妻曾否动粗?”的题目,若两个回答都是“偶尔”时,这两题所得的分数可能是一样的(在 5,4,3,2,1 的问卷分数中大概都是 2 分),虽然在现实生活中,偶尔动粗比偶尔口角严重多了,但在统计分析时,尤其在加总以后,这两个 2 分是同样的意义。

（三）若干题目的总和,可以视为总量表的部分量表

一般的量表都不会只有一二十题,不会只调查受访者对单一事件的态度或意见。一般的量表可能有好几个单元,如吴秀照论文中的“组织融入”这个概念的调查,可分为“对组织的向心力”“对组织未来的关心”“为组织努力的程度”“与组织目标符合的程度”等单元。假设这些单元各有 8 题,那么“组织融入”的调查共有 32 题,但是“对组织的向心力”这 8 题的分数可以视为总量表“组织融入”的部分量表,同样地,“组织融入”这 32 题的量表可以视为“工作态度”的总量表中的部分量表。

（四）总加量表有关态度的测量,经常都是分五种等级来表达

例如,“你对公共场合不准抽烟的规定有何意见”,受访者就在研究者所预备的五个答案“非常同意”“同意”“无意见”“不同意”“非常不同意”中圈一个作答;有的问卷则以是否赞成来区分。有的研究也用总加量表的形式来调查受访者某种行为的频率,如“夫妻口角过吗”,而答案就以“经常”“很少”“偶尔”“从来没有”等方式让受访者圈一个作答。五个层次的答案是非常普遍的方式,有正面、有反面,有极端也有中庸,问题是很多受访者却容易选中间的“没有意见”这个答案来作答。这当中,有的是因为受访者不想动脑筋回答,有的则因为题目太深,问得不具体,受访者不知从何答起,因此拿中间的答案来作答。这种情况虽伤害性不大,但却因此使问卷的辨识力不足,无法清楚地探讨受访者真实的想法,这是五等份答案的缺点。基于此种因素,有的研究者就改以四等份的方式来询问,把中间模棱两可的答案拿掉,倒也不失为聪明的办法。把受访者的意见分一些等份来调查是简单方便的方法,从精确度来看,应该是等份愈多愈好,如分成十一等份或二十五等份,受访者回答的

精确度当然会大幅提高,问题是,很多受访者却没有这么精细的辨别能力,勉强要他把回答切割为十一等份时,受访者往往无从下手,到最后只好以中间的答案来搪塞,这种后果就得不偿失了。

(五)各个问卷题目(items)所得的权数(分数)不是由受访者来决定,而是由研究者主观判断来决定

除了上述由研究者决定受访者就五种答案或四种答案来圈选外,基于研究的假设,还必须根据问题的正向或负向来决定分数是多少。例如,调查婚姻关系时,问“夫妻每天都有十分钟以上的沟通吗”,回答如果是“天天有”是 5 分,“经常有”为 4 分,从来没有为 1 分。问“夫妻口角吗”,若是“天天吵”的话则为 1 分,“经常吵”则为 2 分,从来没有则为 5 分。分数的分配由这些题目的正向或负向来决定。问题是有些研究的正反性质不是很容易分辨,如管教小孩的态度,“当孩子犯了错(偷东西)时,你赞成体罚吗”,这种题目的正负向决定可能就见仁见智了。基本上,这类题正向或负向的决定是由研究的假设而定的。若研究者在研究的结构上以自由主义的反体罚为主轴,那体罚绝对是负向题。反之,若以家长管教的有效性来说,可能会变成正向题也说不定。

(六)分数的计算依研究者主观判断其正负性质后,方式虽有不同,但等距的观念则是一致的

例如,正向的题目可依“非常同意”“同意”“无意见”“不同意”与“非常不同意”等给予 5,4,3,2,1 的计分;若为负向题,则反过来为 1,2,3,4,5。也有 0,1,2,3,4 的方式,或是用−2,−1,0,1,2 的方式亦未曾不可。

二、Thurstone 等距量表

塞斯通式量表(Thurstone scale)是等距量表中使用最为广泛的一种形式。等距量表的制作有别于总加量表,等距量表是在一组经过“考核”的过程后所规划而成的一组题目,其设计方法如下。

(一)研究者依主题设计相关的题目

若是以婚姻满意度为主的研究,研究者就以婚姻满意为主题设计题目,题目的设计或结构过程,最好经过由理论而概念、由概念而变量的过程,以免在结构上产生偏差。

（二）原则上，每一题目中的叙述最好以第一人称、现在式，以及假设的语气设置

例如，“我认为若政府扫黑的工作早五年进行，老百姓对政府的信心势必提高”或“我认为社会工作师资格的取得应该都经过政府的考试过程，不应当有任何的例外”，全部都是在第一人称、现在式以及假设的情况下发问。

（三）当这样一组题目设计完成后，研究者必须把这组题目送请对该主题有研究的专家评鉴

这些专家必须对每一道题目做评鉴，每一个评鉴都分十一个等级，等级的次序从“最不赞成”“无意见”到“最赞成”，这些专家不是针对题目的内容而作答，因为他们不是受访者。专家们是对这些题目将来访问受访者时是否妥当做评估，针对每一个题目都做评估，看是否赞成题目的设计。

（四）计算专家对每一题目的回应，回应愈集中表示专家的意见愈一致

较详细的方法是计算专家们对每题的 Q 值，亦即前四分之一专家与后四分之一的专家之间的差距值，差距值愈小则表示专家的意见愈集中，愈是可用的题目。现因用之者较少，故不再详述。有兴趣者可参阅此书旧版。

（五）选择 Q 值最小的题目构成量表

量表的题次可采用随机的方式，如表 8－2 所示。

表 8－2　Q 值最小题目构成量表的示例

分数	Q 值	原题号	项　　目
1.8	1.3	8	我认为夫妻愈坦诚婚姻愈美满
9.9	1.7	31	我俩容许对方有与朋友共处的时间
9.2	2.8	38	我俩同意共同的目标与期望
8.2	3.0	22	我俩满意于现今的配偶关系
7.9	3.1	24	我们同意子女的管教方式
7.8	3.2	13	个人的信仰生活在婚姻关系中没有冲突
6.9	1.8	30	我们不介意彼此处理生气、沮丧与情绪控制的方式
6.2	3.9	2	我俩能真诚且公开地对谈
4.4	2.7	19	我们懂得如何公平争吵
3.4	1.0	12	我俩很满意于处理彼此差异的方式
3.3	3.3	20	当我们生气时，避免做出伤害对方的事情
2.6	4.5	28	我俩不觉得现今的婚姻生活有压力

(六)只以经过检定的Q值较低的12道题目构成正式量表

但是每题分数多少、Q值多高等信息均不必列在量表上,只有施测者知道即可。

(七)施测时,受访者以上述的题目作答,针对每一题表达赞成与否,每题有不同的分数,再依所赞成题目的分数加总,加总后即为受测者的问卷分数

三、Guttman 累积量表(Guttman scale)

格特曼量表(Guttman scale)是累积量表中使用最为普遍的一种。累积量表主要是根据受访者的回答情况来评定受访者的等级,而不是借着受访者对问卷题目的回答总分来评定。累积量表的问卷题目由具有同一性质的项目所组成。例如,下列一组有关对"黑人是否友善"的态度问卷,有四道题目,而四道题目彼此都有程度上的关联,由受访者回答。

1. 我愿意跟黑人结婚
2. 我愿意跟黑人做邻居
3. 我愿意跟黑人做朋友
4. 我愿意跟黑人交谈

在这四道题目中,我们不仅可以看出对黑人的态度有"层次"上的差异,更有其"次序"的不同。第一题要求最高,第四题要求最低。若受访者同意第一题"我愿意跟黑人结婚"时,可以预测该受访者一定会同意第二、三、四题,即同意跟黑人做邻居、做朋友、交谈,因为第一题最具有限制性。若是受访者同意第一题,却又反对第二、三、四题时,这种态度是前后矛盾的。当受访者只同意第二题时,表示他也同意第三、四题,只不同意第一题。若受访者只同意第三题时,表示第四题也已同意,却仍不同意第一、二题。若受访者只同意第四题时,表示对黑人的态度非常保守,只同意跟黑人交谈而已,绝对不愿与之做朋友、做邻居,更不愿谈到将来与黑人结为夫妻。

设计这类问卷时,每一题都是四小题为一组,四小题的题目安排中都有程度以及次序上的考虑。若第一小题是最为开放的,在没有什么限制的状况

下都同意，那么第二题次之，第四题则会最为保守。对某件事同意了第一题时，表示全部四题都会同意。同意了第二题，代表第三、四题也同意，以此类推。因此若以对婚姻的态度为调查主题时，除了对离婚的态度以外，也可以设计对婚后社交应酬、情感表达、金钱来往、两性关系、性生活的态度等主题，各设计以四小题为一组题目，这些答案的总和可以让我们看出受访者对婚姻或两性关系的态度。

这些题目设计完时，应该先进行试测。假设相关的主题一共设计了二十五题，经过试测，发现其中某些题的答案，试测的受访者几乎都是一样的回应，全部的答案都一致（如都同意第一小题），那就表示题目的设计有问题，区辨力太差，无法借着该组题目探测出受访者的不同意见，研究者就必须把这些题目淘汰、修正或另外再设计。

第五节　问卷的形式

问卷的制作是科学也是艺术，所需注意的，不仅是能否正确地从受访者的回答中探讨研究者所要探讨的主题，还需考虑问卷是否美观、经济，更要注意在日后分析时能否清楚、易于操作。社会进步日新月异，各类问卷层出不穷，有商业性问卷、测验性问卷、社会普查性问卷，以及学术专题的问卷等。有的问卷结构严谨、环环相扣，有的问卷却是潦潦草草，粗糙不堪。有的一两百题，洋洋洒洒；有的却只有三五道题目，简洁扼要。有的问卷采用总加量表；有的为等距量表或累积量表；有的更为简单，没有太多文字的叙述，只是级数表（rating）；有的采用简单的表格，类似清单表（inventories）；也有的用格架表（grids），兹分别介绍如下。

一、级数表（rating）

级数表的制作方法是把所要测量的变量（经常是某种特质的两极状况）分列两端，中间再划分为一些等距的间隔，以表明受访者在这两端特质之间的位置，借此间隔的位置可以看出受访者在该特质中的级数，如图 8－3所示：

请在下列级数表中间的隔数中画“+”来形容你的母亲的特质：

心地温暖的											心地冷漠的
对子女过度保护的											倾向使子女自立的
无私的											自私的
耐心的											性急的
能干的											无能的
欢悦的											沉静的
很像我的											不像我的
粗枝大叶的											敏感的
非常自信的											缺少信心的
很会理家的											不会理家的
很会照料小孩											不会照料小孩
喜欢结交朋友											不喜欢结交朋友

图 8－3　级数表示意图

根据上列的级数表示意图，可以看出级数表所记载的文字不多，只是几个所要调查的特质的正反两面而已，但是根据受访者对该题的勾选位置，研究者一样可以得到很多资料，而且资料内容也相当清楚，不亚于任何以文字叙述为主的问卷。而且研究者可以在两个概念之间设置合适的间隔尺度，以便因主题的难度、受访者的程度和变量的复杂性来决定间隔应该有几个。若难度、程度以及复杂性都高时，则间隔不妨画多点，划分为五格、七格，甚至十一格。若难度、程度、复杂性都低时，划分为三格、五格可能就已足够，否则会难倒受访者，甚至会造成受访者的误解而导致研究偏差，此时就非常可惜了。

二、清单表(inventories)

用清单表的方式来制作问卷也是很有创意的方法。清单表中，左边项目栏所列的问题一目了然，每一题的陈述也比较简单，不用繁文赘字去解释，回答的类别一般只是三五格而已，与一般的问卷差异不大，如有关青少年在学校适应情况的研究(见表 8－3)。

人无远虑，必有近忧，不同的人为不同的事在忧愁。在下列各种事务中，请在适当的格子中圈选你的烦恼状况：

表 8－3　青少年学校适应情况清单表

项　　目	相当烦恼	有些烦恼	很少烦恼
与同学处不来			
社交不好			
担心同学不到我们家来			
常觉得心不在焉			
其他同学嘲笑我			
被取不雅绰号			
被同学轻视			

三、格架表(grids)

若所调查的范围都是一些事务类别或行为动作，而不是较为复杂的态度意见，也不是认知感受与情况，再加上回答可以有多种选择时，那简单的格架表不仅清楚，而且制作也颇为简单，如表 8－4 所示，以医生的处方为例。

表 8－4　格架表示例

	鼻通乐	咳嗽液	喷喉药	糖浆	喉糖	咳嗽药
胸咳						
干咳						
吸烟咳						
严重咳嗽						
普通感冒						
喉痛						
喉干						
失声						
口臭						
失眠						

在这个格架表中，研究者只需列出症状的名称和各种处置的药物，就可以

找出所要的详细资料,这类题目中,把两个栏交叉,其交叉点即所要的资料。

四、一般调查的访问问卷

以黄松林(1994)的硕士论文——《老人养护机构服务提供与被养护者满意度之研究》为例,他调查被供养的老人对养护机构的意见,其问卷系用总加量表的李克特量表来调查,回答分为:非常同意、同意、无意见、不同意与非常不同意五种,如下例(原调查问卷共有26道题目,仅列出11道题目供参考)。

	非常同意	同意	无意见	不同意	非常不同意
1. 这个机构的地点交通便利	□	□	□	□	□
2. 这个机构内外的景观令我满意	□	□	□	□	□
3. 这个机构附近的风景令我感到舒适	□	□	□	□	□
4. 这个机构设在此社区令人满意	□	□	□	□	□
5. 这个机构附近的社区居民是友善的	□	□	□	□	□
6. 整体而言,这个机构的设置地点适当	□	□	□	□	□
7. 我所住的寝室活动空间足够使用	□	□	□	□	□
8. 我所住的寝室采光令我满意	□	□	□	□	□
9. 我所住的寝室通风或空气系统设置良好	□	□	□	□	□
10. 这个机构的安全设施良好	□	□	□	□	□
……					
26. 整体而言,我觉得住在这个机构令我满意	□	□	□	□	□

五、问卷的制作

(一)问卷制作的基本规则

有些问卷是由调查者持有,根据问卷的问题,依序向受访者征询问题,收集资料。有些问卷则由受访者自行阅读,自行作答。当问卷由受访者自行作答时,问卷的设计便充满了挑战,在规划上必须注意以下细节:(1)主题原则,问卷要抓住主题,根据主题来设计问卷,不合主题的不应在问卷中出现,即问卷中只出现你所要调查和分析的内容;(2)吸引原则,问卷愈吸引人愈好,要提高受访者填答的意愿;(3)精简原则,问卷设计应精简有力,否则不仅问卷冗长臃肿,也会使受访者随便作答;(4)礼貌原则,适当地进行介绍、说明,以减少受访者的戒心;(5)预先考虑原则,应在事先就把可能产生的问题予以预防、克服,不能事

到临头时才发现漏洞百出，此时会付出高昂的时间与经济代价，若使研究中断或产生误差则更不划算。

（二）问卷的结构

虽然问卷的内容各有不同，其结构却是大同小异。可分为以下项目来说明。

1. 问卷封面

除非是课堂作业，或是一些较简要的问卷不用封面以外，问卷大体都有封面。封面记载着研究的名称，或加上研究者的机构、指导老师与研究者（学生）的姓名或通信地址等。

2. 问候与说明

为了取得受访者的合作与信赖，问卷开始前应该以信件问候的方式，把该研究的主旨以及请求向受访者表白。该问候与说明可以在第一页（封面的下方）或第二页（问卷正式开始之前）完成。

问卷是要向受访者询问他们对有关主题的看法、意见或态度的，因此访问者丝毫不能大意。Baker（1994）指出一般的调查问卷都有其基本格式。开始总有开头的信函（cover letter），通常为简单的一两段话、礼貌性的简介，来说明研究的目的，并保证应有的隐秘性，必要时可以把指导老师或某个机构主管的签名或推荐函附上，或可减少受访者的疑虑。

3. 单元指示说明

问卷会依主要自变量与因变量的次序，分为不同的单元，而且各单元的作答方式也不尽相同，所以研究者有必要在每个单元之前举例说明，以便受访者可以按提示作答。

说明或指示（instructions）用来告知受访者如何回答，答案应写在哪里；常常会有一两个例题，可使受访者觉得更有把握。是否要用特殊的回答工具，或另有答案卡也应另外说明，若由电脑来阅读处理时，特殊工具的使用说明就更显重要。若某些问题可以跳答或不答（列联式问卷）更要说明清楚，有些问卷甚至用图形箭头来指引。最好也应说明该如何交卷，通常在三个地方说明：附信、问卷结尾或问卷开始的地方。

4. 问卷内容

这是问卷的主体，呈现的方式依问卷的性质不同而有其差异，但是必须有

一个通则以供遵守,大体都是以问卷的主要变量为依据,内容的建构则看是采用何种尺度(scale)为主,但最大的要求是问卷的呈现要合乎逻辑,使读者在受访时也能知道研究的方向,不致如置于云里雾中。

5. 问卷的呈现类型

一样是用问卷来问问题,题目的呈现类型却不一而足。有封闭式(closed-ended)或开放式(open-ended)之分。若是开放式,则问卷中适当的空间安排与线条的规划都要设计,以便让填写者有空白处可填。封闭式的优点是简单、清楚,在资料分析时也较为公正,而且易于分类,使研究者能很快地比较两个受访者的差异,登录也比较方便,受访者通常不会空白或答不知道。但是封闭式问题的缺点是容易产生群式答案,大家都圈选中间那项,研究者便不易区别一些群内的差异。封闭式的问题也会有受访者因不会回答,只好随便圈选勉强作答的情形。封闭式的问卷适合作为概念较清楚时有效收集资料的工具。开放的问卷通常用于探索性的研究,缺点是无法有效归类,所以容易出错,并且容易偏于个人化,各说各话,南辕北辙。所以当我们进行研究而正在规划问卷时,比较保险的方法是先用开放式的问题来当试测,再依试测结果归纳出一般常见的答案类别,正式施测时再根据这些类别做成封闭式的问卷,则此封闭式的问卷就会有较佳的效果。

6. 问卷题次的安排

问卷的安排也有独立性题目与关联性题目(contingency questions)之分。独立性题目每一题都是单一独立的问法,关联性题目就是依据上一道题目所发生的问题再延伸出一些相关的题目,如问“你一天吸几根烟”时,前一题应该是“你有没有吸烟”,再下一题才问“抽烟时的感觉如何”,如此一来,这三个题目事实上是彼此关联的一个主题。

矩阵式的题目(matrix questions)也是问卷题目安排的一种方式。例如,把相关的一组题目都整齐排列,右边的空间则列出一组很相似的答案让受访者选择(如非常赞成、赞成、没意见、不赞成、非常不赞成等)。这种形式的问卷可以让受访者较容易回答,而且节省问卷很多空间,文字的叙述也较为简单,不必每题都重复叙述,但要提防受访者全部给同样的答案。

各种类型的题目中,研究者仍需注意题目的顺序(ordering of questions),把有趣的、引人入胜的、受访者容易回答的题目放在前面,以免受访者抗拒。

7. 问卷文字的安排

问卷的制作虽然都是一道题目接着一道题目，谈不上文章的写作，但是每一题的陈述中，也有不少文字安排的细节需要注意，所费的苦心有时不下于长篇大论。例如，每道题目的用字及其所表达的意义是否简单清楚？对填写者而言，会不会有其他的隐含意义？有时研究者的真正意思无法被受访者完全正确理解（研究生或大学生平日讲话、思考的逻辑等，有时与社会群众有很大出入，因此所设计出来的问卷在内容上难免会与群众脱节，所以要特别注意），受访者常常在不明题意或担心回答可能不符合社会期待的情况下给出错误的资料。原则上要避免负向的问题，而以正向的题目为主，避免让填写者头脑急转弯（一题正，一题负），愈回答愈迷糊（以往的教科书中常常建议问卷中最好正负题各一半，但是经验告诉我们，如此一来反而让填写者搞混了题意，尤其当题意较复杂，或填写者受教育程度较低时，正负题掺杂在一起容易造成填写的错误，所以是否要正负题各一半，要看受访者的状况以及所调查的内容而定）。问题的叙述也要避免一题两问。当问卷完成时，不要忘了检视问卷中是否存在研究者个人的偏见（敏感的人从题目中就可以看出研究者内心对该问题的态度倾向）。问卷问题本身是否有直接或间接为难填写者的情况产生也要注意，不要让填写者产生被捉弄的感觉。

8. 基本资料

基本资料可放在问卷的开头或结尾。当问卷的敏感度不高，受访者不必担心所答的内容是否会泄露时，基本资料置于前头并无不妥。若事涉敏感问题，生怕受访者填写的基本资料会影响受访的真实性时，那就要注意。

9. 回复方式、回邮的规划

问卷的回复方式与回邮规划相当重要，不可忽略。若事关敏感问题，受访者不必具名，而且回邮均已事先预备，无须受访者再去回复时，那受访者迅速回复的概率应该可以提高；反之，受访者的回复率必然会受影响。现代社会中，很多人已经很习惯回答各式各样的问卷，反正花个几十分钟的时间可以协助别人，只要问卷不要太离谱，大体都会把问卷填完。有时“去寄信”反而会是个问题，研究者若是对某个机构内的某些员工进行调查时，那在问卷上记明将于某个时间亲自去收集问卷时，回收的百分比一定不错。此外，有些研究者借着精密的折叠，使受访者不必用信封就可把问卷寄出时，必须注意不要用寄信时的

背面来做受访者回复时的正面(受访者的名字早就昭然若揭,如何能保密?受访者如何能够放心?)。

六、问卷的试测

通常研究者认为好的题目,到真正施测时不见得一定是好的。问卷好坏的最后裁判者是填表者,而不是问卷设计者。比较保险的办法,就是当问卷制作完成后,最好先找一小部分的样本做试测,借试测找出哪些问题常常没答,哪种题目最受欢迎,有关开放性的题目到底效果好不好,矩阵式的题目有没有出现"一组式反应"(群式反应)……找出这些问题后马上改进,以免正式施测时滞塞难行。问卷的试测要回答下列问题。

1. 所设计的问题能不能测到所要测的?
2. 所有的题意、字眼都能被了解吗?
3. 问题能够被理解吗?
4. 封闭式的答案类别合适吗? 能把各式答案都全部涵盖吗?
5. 能够使人愿意回答吗?
6. 问题能够被正确地回答吗?
7. 有没有遗漏的问题? 会不会引发一些不易判断或说明的答案?
8. 有没有研究者本身的偏见或观念夹在其中?

为了有效回答上列问题,试测时最好是由同学(事)来处理,这个同学(事)最好是富有经验者,而且了解其研究目的、类型与假设。若无此类同学(事)可以帮忙时,由资料的使用者,如决策者、其他专业人员或未来的访问者帮忙亦可。试测的样本最好从总体中抽样,如此一来,试测的结果与正式施测时的结果才不会产生太大的差异。

有关问卷结构方面枝枝节节的问题,无法用文字详加叙述。研究者不妨在接到其他人的问卷时,都能以学习者的心态来欣赏、批评并且能建议这个问卷表日后改进的方向。除了不犯问卷设计的错,更应在问卷编排的美观、精简、巧妙上下功夫,使受访者乐于回复你的问题。简而言之,问卷的设计是一门科学,当然也是一门艺术。

第六节　问卷的信度测量

信度(reliability)指研究的信赖度(dependability)与一致性(consistency)。受访者被访问时的回答,或受测时的分数,若再施测一次,或再访问一次时,其结果应该相同。若把同样的测验针对相同性质的团体施测,其结果也应大同小异,因为一个好的研究必须有其信度,就如一支好的枪,被一个受过训练的射击手在相同的状况中使用时,每次的弹着点应该一样。若弹着点每次都南辕北辙,其信度就不佳。一般而言,信度有再测信度、折半信度、复本信度,或分数信度等。兹分别说明如下。

一、再测信度(test-retest reliability)

所谓再测信度系指用同一测验,对同一群受试者前后测验两次,再根据受试者两次测验的分数,计算其相关系数,此系数为该测量的信度系数。例如,针对大学生婚前性行为的态度研究,研究者使用同样的测量中,有30%的大学生表示同意,60%的大学生表示不同意,10%的大学生无意见。然而,在第二次测量中,亦出现同样的结果:30%同意,60%不同意,10%无意见,则研究者可以说此测量是可信赖的、一致性高的,其信度亦佳。反之,若第二次测量中却有40%的大学生同意,50%的大学生表示不同意,10%的大学生无意见,则此测量的可信度就不像前者那样高。

再测信度的计算牵涉到稳定系数(stability coefficient)的问题,即两个测验之间应该有相当的稳定性,其稳定性的高低,可用稳定系数来代表。所以稳定系数也可称为再测信度。事实上,问卷的每一题目的稳定系数也不相同,研究者可从其变化总值中来加以比较。如表8-5所示。

表8-5　A题与B题的稳定系数比较　(%)

A题	同意	不同意	未决定	B题	同意	不同意	未决定
test 1	32	60	8	test 1	45	47	8
test 2	34	61	5	test 2	35	59	6
变化总值	2	1	3	变化总值	10	12	2

从以上的比较中可看出,A 题的变化总值为 6,B 题为 24,所以 A 题的稳定系数较 B 题高。

二、折半信度(split-halfreliability)

顾名思义,折半信度就是把问卷“折”成一半,然后再看受访者在这两半测验上的分数,计算其相关系数,则为其折半信度。这种考验,明显是针对问卷的结构而来,因为好的问卷不仅要“言之有物”,而且在结构上也应系统化。研究者将问卷内的问题分成两组,测试同样的受访者,理论上(如果这个问卷设计得好的话),这两组问卷所得到的分数彼此之间的相关系数应该“高”,它的折半信度应该“强”才合理。

折半信度看起来虽然简单,但是它的奥妙是如何把问卷“折”成两半。不是随便分成两半就可以,必须考虑问卷的架构以及问卷的组合过程以后,才能确定如何折半。若问卷的组合是偶发的,问卷的内容是随机的,那如何能折半?折半以后,如何能彼此相关?所以本章第二节所述问卷的结构与来源就显得非常重要,因为问卷若是经由严密的制作过程,由理论而概念,由概念而变量,或由理论而命题,由命题而假设,则问卷内的每一个问题都不能凭空而来,也不是单独存在,它们应该是环环相扣的。若问卷是由这种过程规划而来,那么把问卷“折半”就不是难题,因为研究者对哪个概念转化成哪几个变量清清楚楚,哪个变量转化成几个问题也一目了然,所以很容易在这种结构下把问题分成奇数题与偶数题,奇数题与偶数题在结构上大同小异。研究者也可以因为问卷结构的严谨,以该变量的题数来折半,不一定非得依奇数与偶数的排列不可(其实依每个变量的设计题目来折半才较合理)……这些概念是处理折半信度时研究者所应注意的。再以例子说明,如表 8 - 6 所示。

表 8 - 6　折半信度示例

老人日常生活的身体活动能力评估
1. 能自我进食
2. 能穿衣、脱衣
3. 能自我照顾(盥洗)
4. 能行走

（续　表）

老人日常生活的身体活动能力评估
5. 能自我上下床
6. 能沐浴
7. 有准时上厕所的习惯
8. 在半年内应该都能自我照顾

研究者可将该八题分成两半，1、3、5、7 为一组，2、4、6、8 为另一组，再计算此两部分得分的相关度，即得折半信度。

三、分数信度（score reliability）

在教育系统里面，评定学生成绩是经常要处理的事，评定学生成绩其实也是一种测量，一个好的测量也应该有其信度，常见的信度评估是“分数信度”。一个真正好的学生最好是每一个老师都说好，不能一些老师说好，另一些老师说差，不要太“见仁见智”，否则信度就差。一个好的答卷，最好是在所有的老师都评定这份考卷的时候都觉得好，而不是张老师说很好，李老师却说很差。分数信度的原理也是如此，研究者可随机抽取相当份数的测验卷，请两位（或两位以上）的评分者分别给分，研究者再计算这两组分数的相关系数，这种相关系数则为分数信度。原理与“好的东西，大家都说好”一样，这是信度原理，虽然不一定就是绝对的，因为总有几个例外，要评估者具备慧眼识英雄的能力才能判定。

四、复本信度（alternate-form reliability）

复本信度的理论与折半信度大同小异，只不过是折半信度的 A 卷和 B 卷，现在改成正本与“复本”（不是“副本”，因为副本与正本要完全一样）；复本是与正本本质相同、结构也相同的问卷，它是另外设计的问卷，它与正本“一致”（但非“一样”），所以理论上两者所测量的结果应该相同，即“信度”应该高，其信度则为复本信度。例如，在期中考时，教授分别要学生填写 A 试卷和 B 试卷，此两份试卷在题数、形式、内容、难度及鉴别度方面皆一致，很自然地，该考生在两份试卷上的得分的相关性要高，才能说这份考卷具有复本信度可言。

第七节　问卷的效度测量

要考验学生的数学能力如何，若老师却出语文方面的试题来考，相信每一个人都会觉得不可思议，评估数学能力当然必须用数学题来考，用语文考题怎么可能考出数学的程度来？“考什么科目就应该考什么科目的内容”，这牵涉到测量的“效度”问题。同样的道理，研究婚姻美满状况时，研究者就必须设计与婚姻美满有关的题目，不能用评估沟通能力的问卷，否则研究就失去了效度。教学中，若老师教的是第七章，却考第八章，也是一样不具效度。还有一些考试，老师把题出得很难，几乎没有人看得懂，如此一来，全班的同学，不管平日成绩好的或是不好的，几乎都是同一种水准，这种考试的效度也不佳。所以一谈到调查、一谈到问卷，我们除了要考虑该调查或该问卷的信度以外，也应考虑其效度。没有效度，再好的信度也没有用。

一、内容效度(content validity)

问卷必须适当地反映应有的内容重点与测量的目标，这是天经地义、理所当然的。问卷的设计因此也必须依据适宜的理论，更应针对调查的主题，以系统的程序和逻辑的方法，详细地把与该主题有关的题目一一规划出来，此为其内容效度。因为强调其研究主题与问卷内容的合理性，故又称为逻辑效度(logical validity)。在学校教育中，考试应该依据课程有关的内容，故亦称课程效度(curricular validity)。

在建立内容效度时，通常需要根据一些专门知识，先把该主题的有关内容与范围详加界定，再从这些经过严格界定的范围中选出一些有代表性的项目与主题(某些考试的场合中，则用抽样的方法)，设计组成该研究的问卷题目。这种借由定义范围与内容，再来规划问卷题目的过程，亦可称之为定义效度(validity by definition)。

内容效度的优点是可以有系统地检视问卷内容的适切性，可以检查问卷是否包括了研究主题应有的内容，各内容之间是否有合适的比例分配。例如，研究者要设计有关“疏离感”的问卷，他必须先定义“疏离感”是什么。若参考了大

量的文献，也咨询了各方的专家，终于把“疏离感”定义为“自我的无力感、人际的孤立感、行为的无规范感，以及对未来的无望感”等四项，那问卷的设计就必须根据这四个“向度”来规划，各个向度之间也必须具有适当的比例分配，整个问卷的结构就可以与主题内容配合，也因此才具有内容效度。

二、实证的证同（empirical confirmation）：实用效度（pragmatic validity）与统计效度（statistical validity）

经过测量，婚姻美满程度高的人，其真实的婚姻状况也应该很美满，否则该调查就不具实用性，也无实证性，根本经不起考验。托福成绩好的人，在美国读书时，他的英文能力应该比托福成绩差的人要好，若是成绩好的人在美国生活时却寸步难行，比成绩差的人还要糟糕，这种事情的解释是：托福考试没有实用效度，要不然就是该生在托福考试时有不诚实的行为，所以才获取高分。理论上，从测量中所获取的资料，应与实际的情形相当，也必须通得过实证的考验。分析这种效度时，相关的效度有同时效度（并行效度）（concurrent validity）、预测效度（predictive validity）与效标效度（criterion-related validity）等。

（一）同时效度（并行效度）（concurrent validity）

经由问卷所得到的调查结果若能与实质的状况相称，表示该问卷具有同时效度或并行效度，否则代表该问卷不能准确反映状况，未能与事实并行。并行效度并不研究受测者未来的成功与否，而是利用业已成功或失败者的分数作为比较的标准。一方面，该问卷应具有足够的辨别效度（discriminate validity），若测量婚姻满意度，则夫妻关系正常者（美满者）的分数应比正在分居中的夫妻高，有这种辨别力，才能称得上测量与事实间具有并行效度。同样地，若要进行学生数学程度的学习成就测验，研究者需从适合的对象中随机抽取学生样本并加以施测，研究者更应将其施测成绩与学生们平日在校的数学成绩加以比较，并计算两者之间的相关系数，两者间应该显著相关。总之有效的测验是能“反映”人类现实生活的测验。

（二）预测效度（predictive validity）

好的测验最好也能预测当事者的“未来”，就如 IQ 高的人，似乎可以预测该人在校的学业成绩会较佳。好的性向测验，最好也真的能预测受试者在其性向的范围中有较杰出的表现。要计算出该问卷的预测效度时，必须运用追踪的方

法,对受试者未来的行为表现做长期继续的观察、考核和记录,然后以累积所得的事实性资料,与当初的测验分数进行相关分析,如此一来,该问卷的预测效度就可以被精确地计算。

(三)效标效度(criterion-related validity)

如何测量"成功"?回答这个问题之前,我们必须先确定什么叫"成功",意思就是先把成功的"效标"找出来。若把成功定义为"收入"多少的效标上,一定比定义在"帮助人"多少的效标上有显著不同。所以在设计问卷时,研究者应该根据主题把适当的"效标"找出来,然后再根据效标,好好地设计相关的题目。因此,研究者在设计问卷时,能否找到适当的效标,该效标是否真的能"显示"研究的主题,便是效标效度的问题。一旦每个概念都有其效标时,研究者的研究结果便可以与效标做比较,若与效标之间的关系密切,则效标效度就高。有时,效标效度也称为实证效度(empirical validity)。

三、建构效度(construct validity)、观念证同度(conceptual confirmation)或引申效度(inferred validity)

一个好的问卷,其结果不仅可以反映现实,其结构也应符合理论。所以问卷的设计应该从一个建构的理论出发,先道出各项关于该理论的各样假设,衍化出各种相关的概念与变量,据之以设计和编制问卷。问卷调查结束后,更应由果求因,以相关、实验、因素分析等方法,考查调查的结果是否符合理论上的结构与见解。从内部结构来看问卷的设计是否合理时,称为建构效度。

若有一"现代智力理论",有三项主要假设:智商随年龄而增长;智商与学业成绩密切相关;智商是相当稳定的。以上三个假设可以称为该理论的结构。研究者若根据上述三个假设编制了智商测验,并把实施测验所得的分数加以分析,如果受试者的受测分数随年龄增大而增加,其智商也在一段时间内保持了相当的稳定性,并发现该智商与学业成就之间确存在其相关性,那么,这些实际的结果,就成了这个测验的"引申效度"的有力证据。

四、问卷调查的效度问题

调查就是要调查所要调查的东西,否则只是白忙一场而已。能够调查到所要调查的、所该调查的,就是效度问题。

（一）问卷的外在效度

为了使调查的样本能有效推论总体，我们必须注意调查问卷的外在效度问题。

1. 研究目的的澄清

调查时，若受访者拒绝回复时，访问者必须澄清研究的目的，而且最好能与受访者的目的一致。否则受访者与调查本身有距离，所给的答案也无法作为研究的参考，所做的推论当然无法正确。了解了研究者的研究目的后，回答时文不对题的情况就可以减少，效度自然会提高。

2. 敏感问题的减少

为避免受访者觉得焦虑或隐私受到侵害，最好事先保证保密而且用匿名的方式来处理资料，使受访者可以放心作答。

3. 避免社会性期待答案的产生

不要让受访者觉得“应该做哪些回答才是正确的答案”，也不要使受访者为讨研究者喜欢，揣摩调查者心意而作答。若是这样，误差在所难免，所以事先必须予以说明，告知回答并无对错之分。

4. 相关问题的处理

因问题多有相关，可能会让受访者以为不必在乎答案的题数，自作聪明，选择自己喜欢的题目作答。因此必须强调每题均要作答的重要性，并且不要问一些不相干的问题，以免加重受访者的负担。

（二）问卷的内在效度

内在效度的主要意义在于使研究能反映所欲探讨的概念程度，所以要避免产生测量误差，其方法如下。

1. 问题要明确：须与研究主题有关，字眼不能暧昧。

2. 字语要简单：不能难以理解。

3. 避免双重的问法：两个问题不要问在一起。

4. 问题要精简：可以让填写者很快读通，马上可以作答。

5. 避免负向的陈述。

6. 要对受访者的知识水准加以考虑，依据受访者的受教育程度与背景设计问卷。

7. 利用试测检定问卷的好坏。

以上这些方法多少都可以增加问卷调查的内在效度,在调查之前多花一些工夫,调查时就可以少受一些挫折与痛苦,事前的防范比事后的弥补重要,在问卷调查时更是如此。

第八节　总加量表的鉴别力(*DP* 值)分析

问卷的好坏不仅要看它的信度与效度,还必须注意它的"鉴别力"(discriminate power, *DP* 值)。一个好的问卷应该有能力分辨得出其程度与好坏。若考试无法把好的学生与差的学生区分出来,那这个考试形同虚设,对学生反而有不良的影响,因为认真也没用,成绩不一定比不认真时好。同样的道理,好的问卷一定要反映出不同的程度、不同的态度与意见,否则不问也罢。借着问卷来分辨受测者的好差、程度、频率或次数时,就是该问卷的鉴别力。

总加量表的鉴别力可以借计算得知。为了使研究时所用的问卷具有足够的鉴别力,在施行正式调查之前,都应先施行"试测";试测的目的是使研究者在正式施测之前,了解哪道题目的用词适当、哪道题目不妥。试测最客观的作用是使研究者知道哪道题目的鉴别力高、哪道较低,并且把鉴别力低的题目在正式施测之前剔除,免得影响整个问卷的效度。鉴别力的计算过程如下。

在总加量表中,根据研究的主题,研究者先设计一组题目,这些题目中每题的分量都是等值的:非常同意 5 分;同意 4 分;没意见 3 分;不同意 2 分;非常不同意 1 分。理论上正向题与负向题要合并使用,有 54321 的题目,也要有 12345 的题目。题数不要太多,通常在 20～30 题。假设研究者找了 25 个受访者当作试测的对象,予以施测后,要做出统计表格,计算这 25 个受访者在这些题目里的得分,若题目一共 75 题,每题的回答方式是 54321 的方式,非常赞成 5 分,非常不赞成 1 分,则 75 道题的分数是从 375 分到 75 分(对 75 题都表示非常赞成是 375 分,75 题都表示非常不赞成为 75 分),当然实际状况中,受访者的回答不会是这个样子。统计表格的格式如表 8－7、表 8－8 所示:

表 8－7　受访者分数统计表

题号	受访者编码(依分数由低到高排列)																								
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	5	4	3	4	5	5	5	3	5	4	4	5	5	5	5	5	5	5	5	4	4	4	5	5	5
2	2	1	2	2	1	4	1	2	2	1	2	2	1	4	1	2	1	2	1	2	5	5	1	2	2
3	3	2	4	3	4	3	……													……					
4	5	3	4	4	2	4	……													……					
5	5	3	1	5	4	4	……													……					
6	5	1	5	3	1	1	……													……					
7	4	1	4	1	4	1	……													……					
8	1	5	4	4	5	5	……													……					
⋮																									
75	……						……													……					
总分	……						……													……					

表 8－8　题目试测的鉴别力计算

题号	最低 25%的平均分	最高 25%的平均分	鉴别力
1	26/6=4.33	27/6=4.5	0.17
2	12/6=2	17/6=2.83	0.83
3	19/6=3.17	24/6=4	0.83
4	22/6=3.67	22/6=3.67	0
5	22/6=3.67	27/6=4.5	0.83
6	16/6=2.67	30/6=5	2.33
7	15/6=2.5	28/6=4.67	2.17
8	19/6=3.17	27/6=4.5	1.33

表 8－7 中，就如一般研究生准备正式调查前的问卷试测一样，25 个试测对象的资料都已经收集，按其分数由低到高排列，从最低的 208 分到最高的 322 分，计算分数低的四分之一(即 25%，共 6 人)对每一题的答案的平均分数，也计算分数高的四分之一(即 25%，也是 6 人)对每一题的答案平均分数，两者相减就是该题的“鉴别力”，分数愈高表示鉴别能力愈高(因为可以让受访者的回答

有较大的差距)。表 8-8 中,可以把鉴别力最差的第四题剔除,因为该题的鉴别力是 0。

实际做问卷的试测时,若理想中的题数是 30 题,则在准备试测的题目时,最好能做到 40 题或 50 题,计算每题的鉴别力后,把鉴别力不高的题目剔除,剩下的就是较佳的题目。若预备的题目太少,有时明明知道有些题的鉴别力不高,却也只能"忍痛接受",其结果往往得不偿失。万一需要再做一次试测,更是劳师动众,不如事先题目预备多一点较划算。

第九节 结 论

问卷设计是研究过程中极其重要的部分,研究的概念如何落实,资料能否被有效地收集,处处都需要在问卷设计上下功夫。好的问卷必须考虑到它的效度与信度,若是在信度和效度上能够清楚交代,调查几乎已成功了一半。然而,在研究的过程中,最重要的"评鉴者"不是专家,也不是种类繁多的信度、效度测验,而是受访者。了解研究对象是谁比什么都重要,研究者必须站在受访者的角度来设计问卷,依受访者的作息来设计问卷,依受访者的语言来进行问卷调查。问卷的制作所依据的是"知识",问卷收集的资料却是地地道道的"人生经历",因此所牵涉的层面相当复杂。但是若调查成功,使我们对问题有更进一步的了解,就算调查的过程千辛万苦,也是值得的。

关键名词

内容效度(content validity):指一种测验使用的题目足以代表课程内容或行为层面的程度。

分数信度(score reliability):研究者可随机抽取相当份数的测验卷,请两位(或两位以上)的评分者分别给分,研究者再计算这两组分数的相关系数,这种相关系数则为分数信度。

再测信度(test-retest reliability):指用同一测验,对同一群受试者前后测验两次,再根据受试者两次测验的分数,计算其相关系数,此系数为该测量的信度

系数。

折半信度(split-half reliability test)：折半信度就是把问卷“折”成一半，然后再看受访者在这两半测验上的分数，计算其相关系数，则为其折半信度。

同时效度(并行效度)(concurrent validity)：指根据测验与目前某种效标相关的程度，用以“预测”现在。

信度(reliability)：指量表或测验结果的稳定性与一致性。

建构效度(construct validity)：指量表或测验内容能测量到理论之概念或特质的程度。

效度(validity)：指一种测量技术是否能真正测量它所要测量的问题。

效标效度(criterion-validity)：用以说明在某种测量工具上所得的分数与外在自变量(效标)之间的关系。

测量(measures)：根据某些准则，把一些对象或事件赋予数字的过程。

量表(scale)：依照一套规则，分派数值于行为现象的过程。例如，以人格测验来衡量个人生活适应的能力。

复本信度(alternate-form reliability)：复本是与正本本质相同、结构也相同的问卷，它是另外设计的问卷，它与正本“一致”(但非“一样”)，所以理论上两者所测量的结果应该相同，其“信度”应该高，其信度则为复本信度。

操作化(operationalizing definition)：把抽象的定义予以实体化，使之可以用测量的方式来衡量此概念时，即谓操作化。

习 题

1. 何谓信度？研究人员是怎样决定一个研究计划是否可信的？试举例说明之。
2. 何谓效度？为何问卷设计要注意效度问题？通常测量效度的方法有哪些？
3. Likert scale 与 Guttman scale 各有何种特质？试以你所感兴趣的主题，各设计五题 Likert scale 与 Guttman scale 的问卷。
4. 调查问卷通常可以用哪些方式加以设计？试举例说明之。
5. 问卷设计应把握哪些原则与重点？试举例加以说明。

第九章 实验研究

第一节　前言
第二节　实验研究的类型
第三节　“实验”的内涵
第四节　实验设计(研究设计)
第五节　一般常见的实验设计
第六节　单案研究设计

摘 要

为了了解问题，用精密的实验方式来处理所欲了解的问题，就是实验研究。实验研究又可以分为两种——实验室研究与实地实验。实验室研究，就是在实验室内所进行的研究。实验室研究把受试者分为两组，即实验组与控制组，其把所要调查的变量完全控制在实验室内。实地实验则把实验对象放在实地的情境里，对其做实验处遇的工作。两者各有其优缺点，我们将于本章做深入探讨。

使研究能顺利达到目标，能回答所要探讨的问题，这就是研究设计的主要用途。一个好的研究设计，应要考虑其研究是否能达到目标，还要考虑其他干扰变量的控制、推论性、信效度等问题。

一般而言，实验设计类型依其准确的程度可分为前实验设计、真实验设计、准实验设计三种，我们将在本章第五节详细说明各种不同的类型，社工同仁应该对这些研究设计有其概念；借着比较这些研究类型，可以使读者朋友们知道进行研究时一方面要有严谨的态度，另一方面更要有敏锐的心智，才能去探讨、去规划。

单个案的研究亦在本章中介绍。量化研究一般都是以量取胜，但实验研究的精神亦可应用到时间系列的个案研究中来。在社工处置的前后，若能有客观的记录，研究者就可以用实验研究的原则来评估社工处遇是否达到效果。

第一节　前　　言

为了了解问题、探讨问题的真相，我们实地去看看，这就是实地观察研究。研究者若设计一些量表、问卷去访问，则为调查研究。但是较为精密的方法，则是用实验的方式来处理。实验研究又可以分为两种——实验室研究与实地实验。所以前面所介绍的实地观察研究、调查研究，加上本章所要讲解的实验室研究与实地实验等四种，可谓是社会科学研究的四大类别。

实验室研究，一言以蔽之，就是在实验室内所进行的研究。实验室研究把受试者分为两组，即实验组与控制组。所要调查的变量完全控制在实验室内，受试者的一举一动都在实验室内研究者的掌控之中，实验项目的实施、如何实施（剂量多少等）都在严谨的规划当中，在实验施行完毕后，比较实验组与控制组之间的差距，这就是典型的实验室研究。

实地实验则把实验对象放在实地的情境里。如 Walster、Cleary 及 Clifford 等对学院入学偏见之研究，调查美国的大专院校是否歧视和反对妇女和黑人的入学申请，这些研究者到学校去，对一些偏见或歧视的问题做实地的实验。这种实验不是在实验室里，而是在真实的情境中。实地实验有实验事项（如让一个人在一组黑人学生团体或在白人学生团体中做出某些经过设计的特殊行为，然后再看白人团体与黑人团体对这种特殊行为有何种差异的反应），只是这些处遇不是在实验室内实施，而是在真实的情境里，所以它有别于实验室实验。实地实验也与调查研究不同。在调查研究中，研究者固然是到实际的情境去调查，但并不做任何的“实验处遇”，他只是依问卷的内容，对样本内的受访者做访问而已，而实地实验却要在实地的情境中做实验处遇。

社会科学中，每一种现象所牵涉的变量都相当复杂，较少有实验研究。但是随着科学的进步，加上各种科际的整合，未来的实验研究可能会大量增加；心理学或教育学里所做的实验研究甚多，社会工作中团体动力方面的探讨也常会应用到实验研究，所以必须对实验研究的方法与应用有所准备。

第二节　实验研究的类型

第一种分类是实验室研究与实地实验(laboratory and field experiments)。实验室研究是一切都在实验室内进行的研究,因为实验的每一个因素都能被精密地控制、测量,研究有了结果后,我们再设法将研究的结果推论至真实世界中,这就是实验室研究。而某些生活情境是研究者无法完全控制的,研究变量无法在实验室的情境中操控,只好在真实世界中进行实验,此为实地实验。如Berkowitz将暴力电影作为自变量,将受试者看完电影后的效果与反应(如攻击性行为是否会增加)作为因变量,然后把受试者随机分派到实验组或控制组,比较这两组在攻击性上是否会产生差异,此即为实地实验研究。

第二种分类则为真实实验和自然实验(true and natural experiments)。真实实验是实验者制造一组情境,此情境里包含着某种自变量(independent variable)的操作,当自变量操作完成后,实验者再测量它的结果(也就是所谓的因变量),实验者期待自变量的变化能使因变量产生差异。总之,真实实验的特性是一个自变量的存在,而此自变量由实验者所操纵,其他的操纵因素还包括实验情境的建立和研究对象于情境中的安排。控制其他因素,操纵自变量,以便看出实验自变量对于整个实验情境所造成的影响,此为真实实验。

当实验者无法操纵任何变量或因素时则称为自然实验。实验的方法是观察一个情况,我们把这个自然的情况当作预测变量(predictor variable),把它和另一个情况(称之为效标变量)(the criterion variable)做相关比较,看这个预测变量对效标变量有何影响力,这种在自然的情况中所做的实验称之为自然实验。例如,Phillips设计了18种类型的媒体暴力形态,统计每一种类型的影片在电视播出后,三天之内全美国的杀人案数是否有受到影响;18种类型的媒体暴力形态是预测变量,播后三天的全美国杀人案数则为效标变量。

实验研究还有一种是假设性实验(hypothetical experiment)。例如,1983年时,美国广播公司(ABC)制作了一部影片《那日以后》(*The Day After*),剧情叙述了核子战争对堪萨斯州一个社区的影响。若我们把此影片本身当成一个刺激(自变量),放映给不同的团体在不同的场合观赏,然后再来观察此影片是

否造成观赏者对核子态度的改变，此即为假设性实验。

总之，实验研究可分为两个主要的类型：实验室研究与实地实验。实验室研究又可称为真实实验，实地实验则可分为自然实验或假设性实验。

第三节　“实验”的内涵

一、实验的特性

既然谈实验研究，我们就应该知道什么是“实验”。所谓“实验”均有如下特质。

Baker(1994)指出，实验一般都包括三大元素(components)：(1) 自变量与因变量(independent and dependent variables)，要说清楚是“什么”在影响“什么”；(2) 实验组与控制组(experiment and control groups)之间的比较，一定要设法使这两组在实验之前相同或相等，所以采用的策略包括随机化(randomization)、比对配合(matching)，设法使研究者与受试者都不知道谁是实验组、谁是控制组(double-blind experiments)，以及注意志愿性样本所造成的问题(problems caused by using volunteer subjects)和霍桑效应(the Hawthorne effects)等问题；(3) 前测后测之间的比较(pretest and posttest)。若再仔细分析，可分为下列七点来说明。

第一，实验有“因果”的特质，实验是要确定时间的优先次序，了解何事发生在先、何事受影响在后，这种断定因果的前后是实验的第一个特性。第二，实验对变量(尤其是自变量)有其控制性，必须设计自变量在哪个时候发生？如何发生？强度多少？以便算出其对因变量所产生的作用，所以自变量也好，因变量也好，最好都在实验者的掌控之中。第三，实验很重视随机分配，使所有的参与者(或事件)有相同机会隶属于实验组或控制组，并使他们有相同的机会可以经验到自变量的层次(level of variable)。第四，实验必须确定最主要的自变量与因变量是什么，连其重要性的程度也可以区分出来。因此何者在研究中是自变量，何者为因变量，甚或何者为中介变量，何者为干扰变量等，在实验研究中都需讲明，最好也能讲明最重要的变量是哪一个，并说出其因由，使读者因此知道

整个实验研究的方向与目标。第五,严谨的实验一次只能完成有两个变量的实验(performing an experiment with two variables);在实验中,在同时间内只能研究两个变量,而将其余变量控制不变。如此一来,我们就可以很清楚地知道变量与变量之间的关系究竟在何种程度与状况。第六,分析原始资料(analyzing primary data),并比较处理过的资料(analyzing treated data),甚至还必须处理不寻常的现象(testing unusual point),以便使研究者探查出实验的结果和未实验之前的差距,更能计算出变量之间所可能的功能性关系(function relationship)。第七,测试是否有其直线关系(testing the straight line finding)。变量与变量之间有不同的关系,如直线关系或曲线关系等,若能找出这些关系模式,实验的结果就一目了然。最后,还得实验剩余变量(experiment with remaining variables)。尽管实验的结果已经发现,但最好也能把尚未证实的其余变量,或一直视之为理所当然的变量再加以实验,以便能把其结果跟之前所得到的结论相比较。若能因之而形成或发现一个关系模式(relating the experiment to a model),那实验的结果就更加圆满了。

二、实验室研究的目的

实验室研究是把所研究的变量放在一个没有"污染"的情况下进行,它希望借着严格的控制与规划,可以使所要研究的自变量与因变量都在研究假设的预期下进行。实验室研究可源于理论,从理论中演绎而来,使理论中的论点能借着严谨的实验来证实。实验室研究也可源自其他研究,借着实验来考核其他研究,并修正其他研究的结果。实验室研究更可提炼理论和假设,从其他实验得到新假设,或从其他非实验却已得到证实的假设中形成相关的假设(to formulate hypotheses related to other experimentally or non-experimentally tested hypotheses),为的是协助建立较为严谨的理论系统。

三、实验室研究的优缺点

实验室研究有其特质,当然亦有其优缺点。优点方面:(1) 它能将研究的情境做一有效的控制,可排除许多影响自变量和因变量的外在因素;(2) 它可较自由地指定和操作自变量,且在研究者变量的操作性定义上,大部分都有很高的明确性,变量的陈述当然都相当清楚,不像调查研究要测量价值观或态度时

那么困难与粗糙；(3) 实验室研究一般都有精密的工具可做测量，误差较少；(4) 可控制操作和测量的环境，将可能的“污染”条件排除。这些优点都是其他研究类型比不上的。实验室研究也有其不可避免的缺陷：(1) 缺乏较有力、较自然的自变量，因为实验的情境都是为了达到研究目的而被创造出来的，人工控制的程度颇为严重，一旦太人工化而不自然时，研究结果就算得到证实，将来在实务的应用上仍然还有一段距离；(2) 太强调实验的精确性和对统计的精细程度，却容易因这种精确性和精细性而与自然事实相差太远(因为事实本身可能还无法如此简单明了)，所以对实验室的精确和精细统计的偏爱是实验研究的缺点；(3) 实验研究的高度人为情境也是一种缺点(太过人为反而偏离了自然)；(4) 实验室实验的过程因为精密，所以内在效度高；但也因为太精密，日后很难复制，所以其外在效度太低。这些都是实验室研究的缺点。

四、实地实验的优缺点

一般而言，实地实验较适合于理论的考验，或去发现实际问题的答案；当理论较为完备，只是缺乏实务的验证时，实地实验的实用性较高。但若理论尚未十分成熟、真伪难辨时，实验室实验的功能就较大，因为它可用来检验理论的真伪。

实地实验因其特质，也有其优缺点。其优点主要有以下三个方面。第一，实地实验比实验室实验更适合于研究复杂的社会。例如，心理的影响、过程及生活情况的改变等，都较可能用实地实验的方式进行。实验室实验虽较精密，但研究较复杂的社会时，却较难被使用。对社会工作的研究而言，实地实验较能有弹性且能广泛地应用到各种不同的问题研究中，其实用性较实验室实验高。第二，变量在实地实验中，比实验室有较强的效能，因为它不是像在实验室内那么人工化，而相关的变量也较为自然，不像实验室实验中的变量都被研究者自以为是地规划。因此，实地实验的效能，通常是强且足以贯穿到分散的实验情境中的。而且，研究情境愈真实，其变量愈有力。第三，实验情境愈实在，愈能提高推论到其他类似情境的效度，所以实地实验的外在效度应比实验室实验要佳。

实地实验的主要缺点是自变量很难操作，不仅实验过程中随时产生的问题很难克服，而且如何随机分配实验组和控制组，以便对自变量有效操作是实地实验研究很难兼顾的问题。大体上，实地实验的缺点是较缺乏准确性，由于外在干扰

变量太多,如何准确地测量因变量也是难题。所以实地实验较自然,却不太严谨。

五、实验的科学性与艺术性

实验有其艺术性及科学性。第一,从艺术性来说,当设计一个实验时,我们必须操作情境,即把真实情境做一种人为复制,以便产生预期的效果,这相当不容易,也是高度的艺术。第二,实验必须让自然科学中的 $S-R$ 关系或是社会科学中自变量、因变量的关系都在实验情境中出现,以便检测假设,此需巧妙的设计。自然科学中,我们不必考虑如何与受试者互动,但在社会科学中,与受试者的互动不仅无法避免,而且相当重要,这种难题的解决需要高超的技巧。第三,实验必须使受试者对实验者的策略保持某种无知,对因变量而言,他们的反应也可能并非最自然的反应,因而实验设计通常会面临伦理的困境,如何克服伦理困境也相当不易。

就实验的科学性来说,传统的科学实验有三大基本的要素:(1)控制自变量,并找出其与因变量之间的确实关系;(2)规划前测、后测,并计算出前后测之间到底有没有显著的差异;(3)分配受测样本至实验组、控制组,并设法使这两组尽量相似,为了达到这个目标,可借着随机抽样法、配对法,或双重盲目实验法(double-blind experiment)(受测者及研究者都不知道自己的组别,而达到较公平的情况)。而整个实验的目标就是在这三种要素之下找出实验到底有没有产生其效果。其科学的程度不言而喻。

实验方法有它的特性。在自然科学中,若要采用"实验"的方法,这个实验方法应该有如下一些特质(Beach & Alvager, 1992)。(1)在实验之前就必须决定什么是相关的因素,实验不能毫无头绪地瞎碰、硬碰,路上随便捡东西,就说是实验的成果与发现。(2)针对两个变量做实验(performing an experiment with two variables)。之所以针对两个变量做实验,是因此能够具体地指出研究的真实进展,可以决定何者为因、何者为果;因如何影响果,在何种状况中会有什么程度的影响。否则万一牵连的变量太多,这些变量彼此又互相产生交互作用(interaction),就算实验结果颇为显著,但到底是哪个变量所带来的作用,或是这些变量交叉融合以后所产生的"突变"而带来的作用,就很难分清楚了。(3)分析原始资料(analyzing primary data)。顾名思义,原始资料就是尚未经过处理(实验)的状况,要对这资料清楚掌握才能知道经过实验以后因变量到底有了多少

变化。(4) 分析实验后的资料(analyzing treated data),以便跟原始资料做比对。(5) 掌握不寻常的案例(treating unusual point),所有的资料都是暴露在同样的处置情境里,为何大部分个案都是我们所能预测的结果,而总有几个个案特别“反常”? 这些反常的案例,必然有它的意义,若仔细研究,可使研究内容更加丰富。(6) 找出线性的趋向(通用的模式)(testing the straight line finding)。(7) 再对剩余变量做实验(experiment with remaining variables)。(8) 把实验结果归纳为一个模式(relating the experimental result to a model),因为实验的结果累积到一定地步时,其模式的建立就如水到渠成一般,自然而然地被建构起来。从以上这些实验的特质来看,我们可以看出实验的“科学性”。

六、因果关系的规则

实验法比其他社会研究方法更重视因果,因为实验法的最终目标是去发现或确定哪一个变量造成了另一个变量的改变。

实验法本身也是一种因果的探索,一方面实验法在追寻一个已知的时间序列(a known time-order sequence),要确定何事发生在前、何事紧随在后。实验法也试图建立刺激与反应之间的关联(an association between the stimulus and the response),了解刺激如何影响反应,分析彼此之间的方向与强度。实验更是寻找一个可能说明所观察到的实验效果之另外的因素,不放过可能的疏忽,也重视各种细节(Baker, 1994)。

社会科学研究者对因果关系常感到不舒服,主要因为实验平常对人类的行为与态度的观察与测量经常不很精确,有时也难免失之于主观;勉强要说出其因果,难免失之于武断。而且社会现象中,两个变量之间的相关性常是不稳定的,随时会受到许多其他因素的影响。况且时间的次序更是很难好好建立。因此,社会科学研究者一般都以关联(associated relationship)来代替因果(causal-effect relationship)。

第四节　实验设计(研究设计)

狭义的研究设计是指当研究者决定要选用“实验方法”时,为求准确、有效、

省时、省力,考虑到合理的抽样,希望有良好的推论性,因此针对研究的变量、情况、特质,在众多的实验方法中,选择一种较合宜的方法,此为实验设计,这也是本节所要特别探讨的,或可称为狭义的研究设计。

广义的研究设计指在有关各种研究的基本类型中,为了要达到研究目标而设计一种方法,以使研究能在准确、省时、省力的状况下顺利进行。例如,在进行研究时,必须确定使用哪一种研究方法,是用实地观察法?还是访问调查法?用文献探讨法?还是实地实验法?这种类型的选择是进行研究时的重要决策,是广义的研究设计。换句话说,研究者在做研究之前,选择了一种方法后(不一定是实验方法)所做的种种设计,可称为广义的研究设计。

研究者在进行研究时,有时虽然没有采用实验研究法,而是采用调查研究法中的访谈方法,但是也可以采用实验研究法中的某些实验设计,作为变量、抽样、样本规划时的参考,对于研究的进行也颇有助益。虽然社会科学中进行一般的研究调查时真正采用严格的实验研究法的概率不高,但是实验研究法的原则与精神,却可以成为任何研究调查的参考。本节就以研究设计为主题,介绍研究设计的主要内容。

一、研究设计的目的

研究设计的主要用途是希望研究能顺利地达到目标,研究能回答所要探讨的问题。更重要的是要确定研究能有实在性(亦即效度)(validity),能探讨该主题所该探讨的内容;对该问题的研究也达到某种程度的客观性(objectivity);调查应该调查的对象;问应该问的问题,使研究具有正确性(accuracy)与经济性(economy),否则就失去了研究设计的意义(Martin,1985;Ray & Ravizza,1985)。

二、好的研究设计的标准

好的研究设计应该符合下列一些标准。

1. 有否回答研究的题目?是否有效地测出假设被接受或排斥?若不能达到这个目标,研究就是文不对题,谈不上是一个合格的研究设计。

2. 变异数(variances)对外加变量或不适当的变量(extraneous inadequate variable)有无控制?

3. 推论性(generalizability)高或低?好的研究设计不是空中楼阁,应该有

极大的推论性，亦即这个好的设计可被其他相同的情境使用。

4. 内在效度与外在效度是否足够？内在效度指研究的过程是否有瑕疵，造成研究误差的因素是否业已排除，研究的结果是否正确。外在效度则指研究的环境是否受到干扰？研究的安排有否产生一些无法预测的偏差，致使研究的延伸程度也产生问题。

当以上问题，研究者的回答都是“positive”（肯定）时，该研究设计就是一个好的研究设计。

三、研究变异数（variances）的控制

从调查与统计的观点来看，研究设计最重要的目标是在研究过程中能控制各种变异数（research design and variance control）。

所谓控制变异数包括三种含义。

第一，设法扩大实验变异数（maximization of experimental variance）。借着研究设计的规划，能使接受实验或处遇的实验组与没有接受实验或处遇的控制组间所产生的差异愈大愈好。让该造成差异的因素，借着实验设计都能顺利地在研究中呈现出来。为了达到这个目标，研究设计必须设法使处遇产生其效果，使之与没有处遇有所差异。

第二，设法控制外加的变异数。研究进行中，很多非研究本身所致的变异数也会干扰研究的进行，这些外加的变异数必须在研究设计中予以排除。为了达到这个目标，我们要设法减少变量（样本本身）所带有的实验以外的附加变异数（to eliminate the variable as a variance）。例如，在实验组、控制组两组当中的样本应该愈相似愈好，按理，在未进行干预之前，其结果应该是一样的。经过处遇以后，若其结果有所不同，研究者也才能理直气壮地说这些差距完全是由实验处遇所带来的，而不是样本本身就已经有所不同。在一般的实验设计中，经常以随机抽样的方法来减少外加变异数，因为借着随机抽样，实验组与控制组的样本就会较相似。若还是不能控制研究所不期待的外加变数，干脆就设法把这些可能产生外加变异数的变量当作真正的自变量来处理（例如，要研究经过团体辅导以后的生活适应力是否较未经过辅导的人要好，实验组与控制组的人应该愈相似愈好，但若研究者担心教育程度可能会影响生活适应，而且若在分派样本至实验组或控制组时，研究者却没有把握能把两组的教育程度摆平，

那就不如把教育程度也当作一个自变量来分析,可用双因子变异数分析的方法来计算教育程度,或教育程度与接受辅导的交叉作用对生活适应所造成的影响)。再不行,干脆就用配对(match subjects)的方法进行实验。

第三,减低误差变异数(minimization of error variance)。要减低误差变异数,就必须在研究过程中设法控制情境,使测量时的误差减到最低限度,否则实验结果便不准确,研究者也无法自信地说是由实验的结果所致。其次是测量所用的量表必须有足够的信度与效度,否则所测量的结果就算显著,也可能与实验无关,这种研究就毫无意义了。所以除了研究本身所重视的主要变量要注意以外,最好也能计算出其他一些无关的变量所造成的变异数,使一切变量所产生的变异数都在掌控之中。

四、影响研究内在效度的一些因素

Campell 及 Stanley(1963)对研究设计的类型提供了非常简洁、明晰的分析,而各种实验类型的好坏,取决于其内在效度及外在效度的优劣。所谓内在效度可从下列几点来说明(Jones,1990;Cook & Campbell,1979)。

(一)影响研究的历史(history)事件

影响研究的历史事件即研究过程中,是否发生一些外在的特殊事件左右了受测者的反应。最出名的例子是对种族偏见的研究。在调查前几天,当地正在放映一个描绘黑人暴力的电影,而且吸引了很多人观赏,当居民对此电影还记忆犹新时调查其对黑人的态度,显而易见,当地居民对黑人的看法一定比未看电影前要差,此电影可称之为影响此研究的历史事件。

(二)个人的身心成熟(maturation)程度

受访者个人的成熟程度以及身心变化也会对研究造成影响。例如,一个研究调查需要很长一段时日(一年半载的)才能完成,虽然都是调查某特定年龄的受访者(如小学六年级学生的身心发展),但较后面时段接受调查的学生,其身心成熟度当然较先前受调查的同年级的同学要好。又如一个研究有前测及后测,若前后测相距的时间甚长,则后测的结果当然比较好,而这种较好的结果不见得一定是处遇产生的效果所致,可能只是因为案主在后测时,身心已经比较成熟所致。因此,若做心理成熟或心理治疗效果的评估时,这种身心成熟的因素不可忽略。

（三）熟悉测验内容（testing）

熟悉测验内容的人在受测时，其结果会比不熟悉其内容的人要佳。例如，出国求学的托福考试，到过补习班上课的人成绩普遍较佳，但不一定是他们的英文程度较好，最主要的原因是补习班业已收集了很多考试题，把各国托福的题目归纳、整理后当作上课的教材，因此补习学生早已熟悉托福考试的全部或部分内容，托福成绩当然较佳。

（四）测量工具有了问题（instrumentation）

实验的结果必须借测量工具来测量，若是测量工具不准确，或是被破坏，所测出来的结果会使整个研究的准确性降低。调查研究中的问卷若是由受访者自行填答，或是实验研究中有关处置是否达到其效果也是由受访者自行填答时，填答的问卷即成了测量的工具，若此问卷由于节省起见已经被用了数次，并已由以前答题的人做过记号时，或是测量工具因某种因素不再准确如初时，这种测量工具所测出来的结果就不准确，研究的内在效度就会降低。

（五）统计回归（statistical regression）

调查受访者对某些事情的意见时，纵使其意见相当极端，但是经由多次访问与调查之后，总会有趋中的现象。群体受访时，也经常会互相左右，在统计上称为回归的趋中现象，这种情形就会影响研究的内在效度。

（六）差异选择（differential selection）

实验计划里一定要把样本分成两组——实验组与控制组。这两组最好愈相似愈好，所以会设法用随机抽样或随机分派的方式，把样本公平地区分为两组。然而，一些研究生在赶论文时（尤其是实验的结果会影响到他的学位时），人性的脆弱本质使然，研究者在不知不觉中，会倾向设法把较好、素质较高的样本放在实验组里。这种差异的选择，会造成实验结果的“显著差异”，但是这种差异是由差异的选择而来，而不是由实验的影响而来。

（七）实验过程中的伤亡问题（experimental mortality）

实验的过程太过冗长，会使受访者疲累，回答时因不耐烦、烦躁而产生不实填答的现象。或整个研究拖延时日太久，原先参与研究的受访者可能搬家，或是本身病亡等因素，都会使研究的完整性大打折扣。

（八）受访者被选择参与研究与本身成熟度之间的交互作用

一样是受访者，哪些人被选择分配在实验组，哪些人在控制组，这已是研究

过程中的重要议题,加上受访者本身也会成长,这两者之间所产生的交互作用,对研究的影响也是非同小可。例如,对×中学三年级的学生做异性态度的调查或实验,这个题目本身已经很容易造成性别或组别之间的竞争,加上×中学三年级阶段对两性关系一定会有不同程度的成熟状况,如此一来,这些相关因素之间的交互作用一定会对研究的结果造成偏差,进而影响效度。

(九) 因果的时间次序问题(causal time-order)

因果的定义似乎很简单,在前的就是因,在后的便是果。社会科学中实际的状况却不是那么单纯。譬如说,是不良的沟通导致了夫妻关系的不和睦?或是因为夫妻关系不和睦导致了夫妻沟通不良?这个简单的问题,若从学术上来说,一定是"假设夫妻两者之间的关系'中立'(关系不好也不坏),而若彼此沟通不良时,会导致夫妻关系的不和睦",用这个方式来确定沟通是因,夫妻关系是果。实际状况中我们知道,夫妻关系不可能是"中立"的,当他们一见面时,可能第一印象已经决定了关系的基础,而在这种关系中,也因此左右了他们彼此之间的沟通动机与意愿,当然也因此影响了沟通的品质,所以什么是因什么是果,是不容易断定的。从事婚姻辅导工作时,面对种种婚姻问题,若要追根溯源的话,常有难断因果关系的困境。

(十) 实验、控制两组之间的相互学习与混淆

明明实验组、控制组各摆一边,各做各的处置,应当桥归桥、路归路,不会有太多的瓜葛。问题是,社会科学的实验与控制两组绝对不会那么听话,不会一切由研究者来摆布。他们会互通信息、互相教导,分享受测时的内容、经验与心得,如此一来,实验组与控制组之间的"纯度"已经大打折扣,实验结果也随之不准确了。

(十一) 对控制组所做的补偿(compensation)

在社会工作的实验或方案计划里,研究者要对实验组进行一项较为合理、先进的关怀方法,对控制组则完全施予和以前一样的"待遇",如此一来,差别待遇明显,研究者会无意识地觉得"愧"对控制组,会在不知不觉中对控制组的成员做或多或少、有形无形的补偿,虽然这些补偿与对实验组的处遇仍然不同,但是在这种情况中所形成的实验差异也就不甚准确了。

(十二) 补偿性的竞争现象(compensatory rivalry)

当研究计划对实验组与控制组有差别处置时,身为控制组的成员可能会有特别的动机与意图,借着自己额外的努力,弥补因不属于实验组的缺憾,与实验组的人产生"竞争"的行为。如此一来,实验的本质就产生了偏差。

（十三）士气低落（demoralization）

与补偿性的竞争相反的现象是控制组成员的“士气低落”，知道自己很不幸是属于控制组，可能是素质较差的一组，既然已经是这一组，那还有什么好争的，不如放弃算了。“补偿性的竞争”反映的是控制组的表现比原先应有的情况要好，“士气低落”反映的则是控制组的表现比原先应有的要差。而实验规划所期待的，则是控制组把其本质不加修饰地表现出来，显而易见地，在这种状况下，实验不可能准确。

五、影响研究外在效度的一些因素

（一）对测验本身的反应或交互作用（the reactive or interaction effect of testing）

测验本身的性质会造成受访者出现一些不属于调查本身所应产生的反应，而这种反应与受访者本身的个人状况再产生交互作用时，调查的结果只会纠缠不清。例如，访问一个刚离婚的暴力受虐女子对美满婚姻的态度时，该女子对婚姻的态度本来就有偏于负向的可能，若问卷的题目又偏偏与她前夫的特质有颇多关联时，这种因对测验本身的反应加上与研究特质间的交互作用，对研究的外在效度会有不利的影响。即使研究的准确度再高，它也是在这种特殊的测验再加上特殊的交互作用下所产生的结果，其外在效度当然不佳，不宜把这种研究结论推展至其他相关研究。

（二）选择偏差与实验变量之间交互作用所产生的影响（interaction effect of selection biases and the experimental variables）

选择偏差会造成受访者的情绪反应、竞争反应，若再加上与实验变量之间的交互作用，研究的外在效度也会受其影响。例如，比较不良青少年与功课优异青少年对金钱价值观的不同时，这种选择性的比较已经很容易造成问题，而研究的主题是对金钱的不同价值观，此时，不良青少年的反应若是太偏激也不足为奇，因为已经牵涉了太多因为实验变量与选择偏差之间的交互作用。因此实验时，所选择的对象愈特别，其外在效度可能就愈差。

（三）对实验安排所产生的反应作用（reactive effects of experimental arrangement）

实验安排中把实验组与控制组放在一起做比较，而且各组都知道彼此是一个竞争的关系，此时两组的反应一定颇不寻常，这种情况会使研究的结果产生

谬误。两组只是为谁输谁赢在计较,而不是实验效果的好坏,这是所谓的"霍桑效应"。可见实验安排本身就会影响其外在效度,安排愈精密,外在效度愈差。

(四)多重处置所带来的搅扰(multiple-treatment interference)

有时研究的实验变量不仅一种,自变量也有很多,不仅各种处置都有其功能,处置与处置间更会有交叉作用的效应产生。更有甚者,有些处置单一实施时可能看不出其效应,但是与其他因素混合实施时,却可能产生无比明显的效应,这就是研究复杂的所在,值得我们注意。处置愈复杂,以后被引用的可能就愈低,外在效度也因此愈差。

总而言之,外在效度代表的是它的概推性和代表性。论及"概推",我们必须考虑其"逻辑性",对事实的猜测必然有其逻辑性存在,而不是随便的推测而已。当我们考虑"外在效度"时,必须常常注意实验设计本身的可推论性、实验结果的可推论性、受试者的可推论性,以及实验情境的可推论性,好的研究是要证明别人在同样的状况下也可以实施,但是当外在效度不佳时,别人在相同问题下所能引用的可能性必定不高,外在效度也就成为其不足之处了。

六、研究设计之内在效度与外在效度的另一种分类

在探讨内在效度与外在效度时,前文以研究的过程以及研究的各种可能的推论性做分析,为使读者充分掌握内在、外在效度的内涵,我们再从对象、实验程序、时间,以及各类的交互作用等方面来分析内在、外在效度的问题。

(一)因研究对象所产生的效度问题

因研究对象而造成效度问题的因素包括选择对象太过特殊、对象在研究过程中亡失、实验组与控制组彼此之间的竞争,或有某组因为较少接受处置因而造成的士气低落等。所以在进行研究时,针对有关研究对象的问题,我们应该注意对象选择的实在性;估计对象在研究期间可能的亡失率;避免实验组与控制组之间做不必要的竞争,也要注意到未受处置那组的心理适应等,以使研究不必因为研究对象的情绪与状况产生与实验无关的变数。

(二)因实验程序所产生的效度问题

有关实验程序的因素包括受访者太熟悉测验内容时的测验因素(testing)。工具本身(instrumentation)在测量方面业已不准确、不灵敏或是产生偏误时,会使研究的过程整体大打折扣。有时"类似的处置"因素也必须克服,研究时只

有实验组才接触自变量，控制组虽然没有接触自变量，但若暴露在与自变量类似的处置或情境下，可能会导致控制组的反应不比实验组差，此时整个分析会有失误。对于控制组的补偿作用也是实验程序中值得注意的事情，因为在一些研究中，当实验组所获得的利益是剥夺自控制组时，便会有设法使控制组获得同样利益的趋势产生，情节虽然动人，但从研究的角度看，却会影响研究的准确性。

（三）因时间所产生的效度问题

实验设计是要证明干预（处置）是否有其功能，证明的方法从实验组与控制组两组之间的差异得出，或是从前测、后测之间的不同来分析。因此，时间对实验而言是相当重要的影响因素。而从效度的角度来看，影响实验准确性的因素有案主本身的成熟因素（maturation）以及历史的因素（history）。人总是不断在成长，后测的结果总是比前测要好。时间愈长，前后之间环境总会有其变化，一些研究者所不能控制的历史事件总会产生，当然也因此使研究的成果多了一些干扰，明明前后之间有显著的差异，但是这些差异不见得一定从实验变量而来，可能只是纯粹实际环境的变化所造成的而已。

（四）其他因素所造成的效度问题

选择何种对象为实验的样本会影响整个研究。最难预测的是样本选择还会与其他因素产生交互作用，使得整个研究变得相当复杂。例如样本选择与成熟之间的交互作用（如选择不同样本至不同组别时，而不同样本本身的成熟速度不同）、样本选择与历史之间的交互作用（如不同的样本各自经历了不同的历史事件）、样本选择与工具之间的交互作用（若测量工具用久失灵，或测量工具被做记号等，以及不同样本用在不同状况的测量工作时）等，都会使研究的准确度降低。其他如统计上的回归问题等也会影响研究的准确度，因此研究者在分析资料时，也应把这些回归的问题考虑进去。

第五节　一般常见的实验设计

Campbell 及 Stanley（1963）对实验研究最大的贡献是他们厘清了各类的实验类型，分析了各类型的特质，并且条分缕析地解释了每个研究类型在内在效度与外在效度上所面临的各种问题。前文叙述了内在效度与外在效度等问题，

本文则转述 Campbell 及 Stanley 的实验设计类型并做简单分析。一般而言,实验类型依其准确的程度可分为:(1) 前实验设计(pre-experimental design),指实验"设计"无科学的严谨性,谈不上是个"实验",只求方便或限于经费、人力而做简单的研究而已;(2) 真实验设计(true experimental design),意即研究设计符合科学实验要求,是真正的实验设计;(3) 准实验设计(quasi-experimental design),当某些情境不能用真实的实验方法来控制变异量时,我们就利用准实验设计的方法。这种方法虽然不能完全符合实验设计中的严格要求,但是大体已抓住了科学的精神与实验应有的态度,只是碍于现实,无法把整个实验的精神完全落实(因为人类行为无法像实验室中的其他变量一样能被严谨控制)。准实验法中,研究者必须了解有哪些特殊的变量是无法控制的,并对该控制不了的变量特别注意,以便使之更精确。社会工作的研究中,对案主完全控制是不大可能的,因此所有的实验均不可能完美,但对于实验的缺陷我们应该了解、分析、承认,以便能使解释更客观。准实验设计包括单组实验设计、五个重复考验的一般形态、分区处理、事后回溯、拉丁方格等。

一、前实验设计

(一) 单组末测设计(one-shot case study)

$$X \qquad\qquad O$$

说明:X 代表一种实验或处遇,O 代表观察或结果。就如要评估一支枪到底准不准时,我们把枪拿起,往标的一打,就以有没有打中红心来评估此枪到底准不准一样。万一打中可能是瞎猫碰上死老鼠,万一没打中可能是其他外力的干扰,我们不能只凭一次的"实验"就断定该枪的好坏,所以这种方法是单枪的个案研究(one-shot case study),谈不上是严谨的实验。社会工作的研究中,要断定一种处遇是否有其效果,我们也不能随便"实验",只看那些经过"实验"的人效果的好坏,就断定该实验是否有效。

这种研究之所以只能称为前实验设计,是因为没有另外一组可比较,因此无法肯定 X 的作用,O 的结果可能没有 X 时也一样可以产生,O 只是推测的结果而已。这个方法的缺点是没有控制组的比较,所以在推论上也不科学。弥补的方法是收集特别的细节,小心地观察,不然就做好标准化的测验。

（二）单组前测、后测设计（one-group pretest-posttest design）

$$O1 \qquad X \qquad O2$$

说明：此设计不像第一种设计，只在处遇完后硬说 X 有效，它是在处遇之前先做前测观察 $O1$，然后比较与后测观察 $O2$ 之间的差异是否显著。此设计的特点是可以比较 $O1$ 与 $O2$ 的不同，但缺点是两者之间的差距除了 X 这个处遇以外，还有其他不少因素。例如，工具（measurement）的问题会使受访者的敏感度增高，后测当然会比前测要好，其他如历史（history）的因素、成熟（maturation）的因素，加上统计回归（regression effect）的问题（后测的分数会有归回平均数的趋中现象产生），若无控制组与之比较，势必无法客观。

（三）双组、无控制设计（two-group，no control）或小组静态比较（static-group comparison）

$$\begin{matrix} X & O1 & \text{或} & X & O1 \\ & O2 & & & O2 \end{matrix}$$

说明：此设计的特质是多了一个组别可以与实验组来比较，如上所示的 $O2$，比较的时间可以同时，也可以把时间错开，一前一后。

这个方法的最大困扰是与实验组对照的这一组不一定就是所谓的控制组（姑且称为比较组），因为这组的形成并不与实验组的形成有相同的手续或过程可以让我们确定在实验之前两组完全相同。因此，若两组有所不同，可能是本来就已存在的差异，而不是实验所造成的。

二、真实验设计（true experimental design）

真实验设计的结构要比前实验设计严谨得多。最主要的是它采用随机抽样的方法，而且还有前测、后测，以及实验组和控制组的比较。主要的类型如下。

（一）前后控制组的设计（pretest-posttest control group design）

$$\begin{matrix} R & O1 & X & O2 \\ R & O3 & & O4 \end{matrix}$$

说明：如上所示，此类实验设计的主要特质是样本被随机分派至实验组或控制组，而且观察比较中，除了有实验组与控制组的比较外，两组都还有前后测

的比较。因是以随机分派法将受试者分成两组,所以这两组在理论上能力应该是一样的。在进行实验时,两组均接受前测,其目的是要检测 $O1$ 与 $O3$ 是否不同,而在实验处理完时,再进行后测,比较其 $O2$ 与 $O4$ 之间的差异。这种设计最大的优点是可以使"历史""成熟""测验""工具"及"回归"等五个问题得以克服,因为即使有这些内在效度的困扰,也因为两组都面临到了,所以在比较上可以"扯平",并不致造成误差。也因为是用随机分派法,所以"选择"、受试者的"流失",与"成熟"的"交互作用"相抵消,因此在内在效度上可以说相当完美。但是在外在效度的考量上,这种设计因为强调了前测与后测,难免会牵涉到测验情境的可推论性,所以外在效度多少会受其影响。

此法虽好,但在分析其资料时,仍不可大意。一般最普遍的分析方法是用 t 检定来看前后测分数的比较,再加上实验组、控制组的比较(即 $O1$ 比 $O2$;$O1$ 比 $O3$;$O2$ 比 $O4$)。当以这种方式进行比较时,有如下的缺点:此时所算出来的变异数比原始的分数(gain scores)不可靠,本质上实验组与控制组的平均数有差异,但有时在检定上却看不出其差异,因此,这类设计最好用变异数分析(ANOVA)才是较正确的方法。共变数(covariance)指足以干扰实验结果的自变量彼此之间的相关性,这在分析上是个难题,所以在分派样本时,一定要用配对法或随机原则来处理,以控制影响因变量和自变量之间的变异关系。若无法做到随机分派时,只好设法把共变数对因变量的影响力除去,再看实验处理组之间的平均数差异是否依然达到显著水准。

(二)所罗门四组比较设计(Solomon four-group design)

R	$O1$	X	$O2$
R	$O3$		$O4$
R		X	$O5$
R			$O6$

说明:上一种设计虽较常被使用,但所罗门四组的实验设计却具有较高的评价,不仅对于内在效度的考虑相当严谨,外在效度的考量也较上一种为优,因为从 $O6$ 与其他组的比较,我们可以算出到底样本对测验所累积的反应有多严重(如 $O6$ 优于 $O5$;$O5$ 优于 $O4$;$O4$ 优于 $O2$)。借 $O2$ 比 $O4$;$O5$ 比 $O6$ 及 $O3$ 比 $O5$ 的分析可以看出处遇(X)的功能,借 $O6$、$O1$、$O3$ 的比较,可以综合出成熟以

及历史的效应到底有多少。换言之，此法主要的特点在于将“有无前测”此一变量纳入实验设计中，将其所造成的变异数部分从总变异量中排除掉，以看出所产生的影响是否显著。从研究的严谨度来看，这个类型的设计相当严谨，但在实际的研究环境中若要依此来实施可能颇为烦琐。

（三）双组比较仅后测设计（post-test only control group design）

$$\begin{array}{ccc} R & X & O1 \\ R & & O2 \end{array}$$

说明：当所罗门四组比较设计的后面两组不易找到前测工具，或前测可能引起副作用不方便行使时，可用此法来考验所得的资料。可使用考验独立样本平均数差异的 t 检定，即可确定其实验效果是否显著，但若有三个或以上的实验处理时，则用变异数分析（ANOVA）来加以考验。若是我们确定两组在实验前应该是平等的，则使用此法，应是较所罗门四组比较设计来得经济。

此法与双组、无控制设计或小组静态比较设计略同，但多了随机抽样 R。此法控制了处置的主要作用及所产生的互动，但并未测量它们到底有多少。若随机的过程没有明显的问题，则此设计优于前后控制组设计，因为前后控制组设计的优点此设计都有，而且看起来简明扼要。

以上三种真实验设计都非常严谨，但仍需考量其他事项。

1. 多因子实验设计（factorial design）

多因子实验设计是指实验研究者同时观察两个或两个以上的自变量对一个因变量的影响，而且也考虑了自变量之间交互作用效果的实验设计。例如，“研究台湾地区不同教育程度与宗教信仰对妇女生育率的影响”。又如，“不同辅导方式（自变量）以及付费意愿（自变量）对辅导效果（因变量）的影响”，辅导方式与付费意愿各有各的主要影响力（main effect），但两者之间还存在一些交互作用（interactioneffect），此种交互作用在多因子设计中经常发生。这种交互作用可以分为三类。

（1）没有交互作用的存在，如不同的辅导方式与付费意愿无关，这两个自变量对因变量辅导效果或许都有作用，但彼此之间没有关联。

（2）两者之间有次序性交互作用。即两者之间有互动的现象，但这种互动是单调型的，第一自变量对第二自变量产生方向性的影响。如接受个案工作辅导者比接受团体工作辅导者较愿意付费，而且接受个案工作辅导的次数愈多，

付费的意愿愈高。

(3) 无次序性交互作用。两个自变量之间呈现一个极端的交互作用形式，付费的意愿甚为复杂，接受个案工作辅导少于三次者付费的意愿虽高，但是超过七次时，付费的意愿却反较接受团体工作辅导者为差，此种形态即为无次序性交互作用，指辅导方式与收费意愿间并不是单一方向的交互作用。

2. 隔宿分类或阶层分类(nest classification)

若进行某一研究，其主题是“不同管理方式在不同感化机构的辅导成效研究”。不同管理方式为自变量 A，不同机构为自变量 B，辅导成效则为因变量 C。自变量不同管理方式 A 有一些“因子”(level)，如人性化管理($A1$)、军事化管理($A2$)。另一自变量不同机构 B 亦有一些因子，如少年观护所($B1$)、少年感化院($B2$)、少年监狱($B3$)、成人盗窃监狱($B4$)、成人烟毒监狱($B5$)、成人凶杀监狱($B6$)。辅导成效在经过测量后用数字来表达，数字愈多表示效果愈好。若研究过程中，我们对于少年犯罪者的机构($B1$—$B3$)都用人性化管理($A1$)来处理；成人犯罪者的机构($B4$—$B6$)都用军事化管理($A2$)来处理，当作我们的实验研究，以便作为日后狱政处理的参考，此种研究设计即为隔宿分类设计或阶层分类设计。如表 9－1 所示。

表 9－1　隔宿分类设计示例

	不同管理机构					
	$B1$	$B2$	$B3$	$B4$	$B5$	$B6$
$A1$ 人性化管理	7	10	9			
	6	6	10			
	5	7	8			
	8	9	11			
$A2$ 军事化管理				5	9	3
				9	4	7
				4	8	5
				3	7	3

3. 变量中的因子(levels)

每一个变量都包含一些因子，这些因子的形式如下：有限模式(finite)、随机模式(random)、固定模式(fixed)、混合模式(mixed)，兹说明如下。

(1) 有限模式(finite model)

某些实验变量的因子是有限的,如人种可分为黄、白、黑、棕、红等五种;或教育程度分为小学、初中、高中、大学、硕士、博士等,或再加上不识字和博士及以上。

(2) 固定模式(fixed model)

在各式各样的辅导模式当中,我们只选两个处遇模式(因子),又如在辅导次数中当然可以从0至30次以上,但只采用三种:5次以下、5～9次、10次以上,此称为固定模式。

(3) 随机模式(random model)

如大学社会工作系二年级的同学共85名,在实习的课程中被安排至不同的机构参观,共参观了四所儿童福利机构、三所家庭婚姻辅导机构。然后要看他们对所参观的不同机构的评价,但是碍于时间,我们不做全部学生对所有参观机构的评估,则从85名学生中随机抽20名,四所儿童福利机构中随机抽两所,三个家庭婚姻辅导机构中随机抽两个来做访问评估,这种模式可称为随机模式。

(4) 混合模式(mixed model)

某因子采用固定模式,另一因子采用随机模式,此称为混合模式。如从某地区所有小学中随机抽选十所,比较启发式教学与填鸭式教学所造成的差异(事实上教学方式不仅启发式与填鸭式两种而已,但是我们只固定为这两种)。两种变量的因子采用不同的方式取舍时,称为混合模式。

4. 从不同的后测时间对实验结果进行测量(testing for effect extend in time)

采取了一种新的措施,要了解该新措施是否有效,牵涉到在什么时候来测量,这种"后测"的时间选择在实验设计中是重要的因素,不可大意。后测时间的决定应包括在研究设计中,Hovlands group 的设计是个很好的范例:

$$
\begin{array}{llll}
R & O1 & X & O2 \\
R & O3 & & O4 \\
R & O5 & X & \cdots\cdots O6 \\
R & O7 & & \cdots\cdots O8
\end{array}
$$

当 $O2$ 比 $O4$ 好,而 $O6$ 比 $O8$ 好,而且比 $O1$、$O3$、$O5$、$O7$ 等都要好时,我们才能肯定地说,处遇 X 的确已经达到它的效果。更重要地,为了测出处遇是不是在一段时间以后才产生其效果的,我们也必须证明 $O6$ 是不是比 $O2$ 好,而

$O8$ 却不见得比 $O4$ 好时,才能显示出 X 的真正功能,万一 $O6$ 比 $O2$ 好,而 $O8$ 也比 $O4$ 好时,我们会发现它们的差异可能来自 X 的因素外,也可能是成长因素或历史因素所造成的。所以在这种实验设计中,我们必须常常注意 X 与历史的交互作用,避免产生不必要的干扰。

5. **处遇方法的概推性:如何使处遇在执行上有变化**(generalizing to other Xs:variability in execution of X)

科学的实验是在实验被证实以后,能够在别的地方、在同样的状况下被引用以后依然证明它的效果,这是处遇方法的概推性。为了使 X 有确定的概推性,研究者必须泛化实验的处理,借着重复的实验来稳定实验效果的解释。此外,研究者也必须好好控制实验设计以外的无关变量,预防它们在实验过程中也对 X 产生了影响。若是控制外来的变量有其困难,则不妨把 X 放入不同的类组中去实验,则处遇方法推论的正确性必可提高。

6. **处遇方法的概推性:不断使处遇更新并与控制组比较**(generalizing to others Xs:sequential refinement of X and novel control group)

“不断改良”是一种科学负责的态度,在研究法的领域中是很重要的,当处遇被证明有效后,对此处遇改良的意义是把处遇概念化为几个变量,并且不断地修正这些变量,使其更精确,详细说明实验处理的过程,另外并发展新的控制组,使自变量对因变量所产生的效果能够在许多控制组中被发现。

7. **概推至其他的观察**(generalizing to other Os)

任何的实验,自变量(X)必然有其他的影响性,但此影响性再精确也会有杂质,一定会有些无关的变量所产生的变异数。所以若要从一个观察推论到其他观察时,我们应事先了解整个研究设计的内在效度是否足够应用到其他的环境时会有什么样的改变……若是内在效度有些困难仍无法克服,那就尽量重复测量一些不同情境下的结果,若我们所要的内容能出现在每一测量时,则其推论性就更高。

三、准实验设计(quasi-experimental design)

如前所述,当某些情境不能用真实的实验方法来控制变异量时,可利用准实验设计的方法,这种方法虽然不能完全符合实验设计中的严格要求,但是大体已抓住了科学的精神与实验应有的态度,只是碍于现实,无法把整个实验的精神完全落实,此即为准实验设计。

（一）时间系列(time sequential test)

$$O1 \quad O2 \quad O3 \quad O4 \quad X \quad O5 \quad O6 \quad O7 \quad O8$$

时间系列设计是对某些团体或个人所进行周期性测量的过程，将之呈现在时间系列中，以观察实验所产生的改变状况。时间系列设计和前实验设计第二种单组前后测的设计在逻辑上是相同的，因为它表面上同样缺乏控制组，虽然都使用了前测与后测。

在时间系列的实验中，若产生了改变，研究者则必须了解这些改变可能有许多对立的解释(alternate explanation)，这种解释也可能与处遇(X)的效果相左，因为实验的时间愈长，愈容易产生干扰的变量(extraneous variables)，会影响整个研究的效度。

时间系列研究法在内在效度上无法克服“历史”“工具”“施测”的问题。所以有关“历史”方面，应尽可能做到实验的独立(experimental isolation)，使其不受外界事物的干扰，不然就尽可能系统地记下各种实验外的种种刺激，并了解在实验过程中有哪些事实可能严重干扰实验的效果。

此外，不妨更加延长且周延地观察，以便了解是否有其循环性，方便找出其恒常的状态。弥补时间序列研究法缺陷的最好方法是设计一个控制组去比较，这种方法可以弥补很多实验设计上的不足。此外，时间系列研究进行时，不要随便修改测量的设备，实验者需固定设备，否则测量的结果必不准确。施行测量的间隔时间也不要太长，否则会增加研究的“历史”事件，影响整个研究的内在效度。外在效度方面，实验的结果只能推论到重复测验的族群，所以有关研究对象特征的描写非常重要，若研究对象不同，时间系列的推论性就有更高的冒险性。复制(replication)也是一种方法，由不同的研究者在不同环境中重复进行测验，可使结论愈来愈准确。

分析此时间系列的资料时，不要只把实验处理的前一次($O4$)和后一次($O5$)做比较，应该更仔细地观察每一次的变化，注意中途是否有较为反常的变化产生，此外，变化幅度(斜坡，slope)的处理也是重点，若斜坡的幅度反常，则在实验后更有解释的余地。

（二）相等时间样本设计(equivalent time sample design)

对一组受试者抽取两个相等的时间样本(time sample)，在其中一个时间

样本里出现实验变量($X1$),在另一个时间样本里不出现实验变量($X0$)的实验设计称为相等时间样本设计。例如,研究者要知道社会工作者与案主面谈时,若有助理在旁边录音做记录会不会影响案主的情绪,或会不会降低案主对被辅导的满意度。假设案主每周接受工作者辅导一次,研究者让助理每礼拜间隔出现一次,来看案主的满意度是否有受到影响,此即为相等时间样本设计。

$$X1 \quad X0 \quad X1 \quad X0 \quad \cdots\cdots$$

此法在内在效度上缺点较少,因为可以借着一个时间与另外一个时间的变化来计算误差到底有多少。但若考量外在效度时,该注意的地方甚多。首先,实验安排的反作用效果必须克服,在这种实验设计中,受试者很容易知道自己正在接受实验,因此实验结果不能推论到没有这种反作用效果的实际群体上。其次,选择偏差与实验变量的交互作用也会影响外在效度,不同的样本对实验有不同的反应与看法,若样本的选择不客观,其与实验变量的交互作用当然也会产生偏差。而且重复实验处理的干扰,虽在误差上或许可以计算出来,但是实验结果却很难应用到正常的群体上(总不能要求正常的群体也都像实验样本一样接受连续的实验处遇)。此种研究设计可以用在只有一个受试者的情形中,但是如果有许多受试者时,所做的实验最好一次或两次即可,否则时间的抽样误差会很大,有太多的因素不是实验设计所能控制的,即使在效果上非常显著,推论上也应谨慎,有可能是累积的实验所带来的效果,而不是 $X1$ 或 $X0$ 的差异。

拉丁方格的设计可以使可能发生的误差大致平衡,而终能互相抵消。在相等时间样本设计中可以参考该设计,如表 9-2 所示。

表 9-2 拉丁方格设计示例

组 别	实 验 处 理			
	X_1	X_2	X_3	X_4
g_1	t_1y	t_2y	t_3y	t_4y
g_2	t_3y	t_1y	t_4y	t_2y
g_3	t_2y	t_4y	t_1y	t_3y
g_4	t_4y	t_3y	t_2y	t_1y

这种实验设计中包含着三个主要的处遇变量：X（处遇）、g（组别）、t（时间）。y 则是处遇的效果。研究者利用四组未经随机分派、维持原来团体形式（如原来班级）的受试者，每组在前后四个时机里，重复接受四种不同的实验处理。因此，在四个纵行的总和中，可以比较四个处遇的差异；四个横行的总和中，也可以看出组别之间的差异。把四个时机的总和再互相比较，亦可以看出不同时间的不同效果。但因所选择的四组受试者是不同质的，所以即使"实验处理"水准间的差异达到显著水准，仍有可能是"组别"和"时机"两个自变量之间的交互作用所造成的。所以有关交互作用的处理，Campbell及Stanley建议：表方格内各细格的分数以 $Mgt-Mg-Mt+M$……来表示，亦即算出总交互作用的总数后，把组别和时机的主要效果予以排除。如此处理之后，若四组都是某个实验处理的效果最强，则研究者更有信心说实验处理之间有显著性差异存在，而且没有组别和时机的因素混淆在内。

（三）相等质量样本设计（equivalent materials samples design）

等量样本设计是以相等的样本来比较是否是因"实验变量的处理"而造成的差异，通常相等的时间也被包括在内。如下所示。

$$M\mathrm{a}X1\ O \qquad M\mathrm{b}X0\ O \qquad M\mathrm{c}X1\ O \qquad M\mathrm{d}X0\ O \qquad \cdots\cdots$$

在这种实验设计中，$M_{\mathrm{a}}=M_{\mathrm{b}}=M_{\mathrm{c}}=M_{\mathrm{d}}$[例如，安排资优班的学生分为四组（假设他们为同等质量），X_1 为实验处遇（上课时由分派学生自行讲解），X_0 为未进行处遇（主要由老师讲解）]。然后在相同的时间段后观察是否有不同的结果。

（四）非相等控制组设计（nonequivalent control group design）

$$\begin{array}{ccc} O & X & O \\ \hline O & & O \end{array}$$

系由未经随机抽样的两组不相等的样本所组成。实验组与控制组均有前后测，常用于无法打破团体分子、无法采用随机或配对的团体。此种研究设计可控制"历史""成熟""测验""工具"等因素的干扰，但是无法控制"选择"与"成熟"的交互作用，也无法克服回归的问题。外在效度方面，前测会增加受试者的敏感度，会产生测验与 X 的交互作用；选择的样本在不同组别时，也会与 X 产生交互作用，都会影响外在效度。若研究者倾向于选择某一类

型的受试者时,会影响推论。实验安排的反作用效应(霍桑效应)也应考虑。

(五)平衡对抗设计(轮换实验设计)[counterbalanced design (rotation experiments)]

此法采用拉丁方格(见表9-3)的安排,以使可能发生的误差大致平衡而相互抵消。X 为实验处遇,t 为不同时机,g 为组别,O 则为观察。研究者可利用四组未经随机分配、维持原来团体形式(如原来班级)的受试者,每组在前后四个时机中,重复接受四种不同的实验处理,可比较不同 group 的所有效果。可利用四个纵行总和,考验"实验处理"之间的差异显著性。计算四个横列的总和,考验"组别"之间的差异。可利用四个各时机总和,来考验不同时间差异是否达到显著水平。再用 $Mgt-Mg-Mt+M$ 的方法,把"组别"和"时机"的主要效果予以排除,只剩下所需要的实验效果。

表9-3 拉丁方格示例

	X_1	X_2	X_3	X_4
group A	$t1O$	$t2O$	$t3O$	$t4O$
group B	$t3O$	$t1O$	$t4O$	$t2O$
group C	$t2O$	$t4O$	$t1O$	$t3O$
group D	$t4O$	$t3O$	$t2O$	$t1O$

四、实验的道德考量

实验方法固然严谨,其科学程度也较高,但是牵涉的却是更多的道德考量问题。从参与者的随机分配来说(random assignment of participant),在真实的社会环境中要做到完全的随机分派是不太可能的,有时为了达到研究目的,在对象的选择上会有限制与偏颇。从处理变量(manipulating variables)来看,要使研究做得好,就必须处理好变量,但处理得太好时,事实上就等于在控制一个人的生活,这样就产生伦理问题。除非我们只用较没有威胁性的实验来测试,当然这样一来,研究又受限了。若要减少参与者的偏差,最好让他们在不知情状况中受试(blind procedures),或不让他们知道研究假设,这样或许可以减少不必要的测验交互作用,或是霍桑效应。问题是研

究伦理上我们却必须给他们足够的信息，使受试者决定到底要不要参加实验，如此一来，两难之间就无法有完美的选择了。因此，若是研究的方式遭到参与者的反对，我们必须再思索其他的方式，而最完美的方式在哪里？我们又能掌握到什么地步？困难就因此产生了。研究结果的展现与问题的解释也是一大难题：我们必须避免过度的一般化（over-generalization），应该只强调被证明出来的部分，因此，在这种情况下，谦虚与诚实的态度就非常重要了。

实验设计或研究设计的钻研，与其说是“如何使用”的问题，不如说是“研究态度”的问题。以上各种实验设计可谓林林总总，社会工作同仁真正能使用的机会并不大，但是社工同仁应对这些研究设计有其概念。通过比较这些研究类型，可以知道，进行研究时一方面要有严谨的态度，另一方面也要有敏锐的心智，毕竟，很多问题需要去探讨，很多事更需要去创造、去规划。

第六节　单案研究设计

一、前言

一般而言，实验研究的使用都是在量化研究中，借着随机、借着实验组与控制组的比较，也借着前后测之间的差距，来证明实验处置的确产生了作用。在这种研究设计中，所用来比较的样本经常都是数目繁多的，以便可以经由统计的分析来确定研究假设是要接纳还是要拒绝（单案设计在社会工作专业领域中具很高的实用价值，社工同仁若惧于上述种种实验设计，此单案研究设计单元或可稍加重视，把此研究设计应用到实务工作里。）

直觉上我们会以为单案研究大概是质性研究中的个案研究，然而亦有人把实验研究的精神应用到单案的研究中。从哲学的背景来分析，单案研究有强烈的行为主义观点，即在处理个案时，处置的目标以案主可以观察、可以计算的问题行为为主，如打架的次数、旷课的时数等，研究者先计算案主的问题行为基准线（即未干预之前的问题行为次数），在经过干预或处置后，来比较干预前后之间的差异，此为单案研究设计的主要内涵。

二、单案研究的定义

单案研究基本上是偏向准实验设计的研究方法(不是在实验室中来进行,而是在实地的情境中以实验方法的原则来处理),它对方案处置的前后做重复而多次的比较评量,被研究的对象本身是实验组,也是控制组。单案研究主要是评量个案前后的变化,用来确定服务过程中改善的程度有多少。在社会工作的实务领域中,单案研究有很大的使用空间。由于要算出处置前后的差异显著不显著,所以单案研究里的观察项目(处置的目标)一定要能被操作化为可测量的项目才可以。

三、单案研究的类型

依测量(观察)的时间点以及测量(观察)的次数,单案研究可以分成如下一些类型。

(一) $A-B$

此为单纯的前测后测型。如图 9-1 所示。

A 为处置前的观察,B 为处置后的观察,X 为处置。纵坐标为处置的结果,愈高表示效果愈佳,愈低则表示效果愈差。

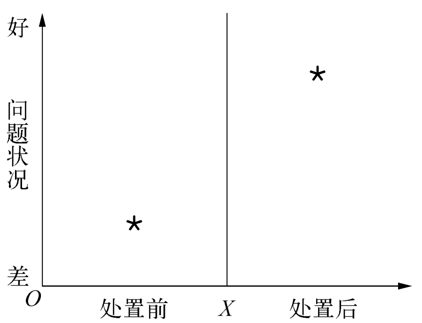


图 9-1 单次观察后的处遇结果

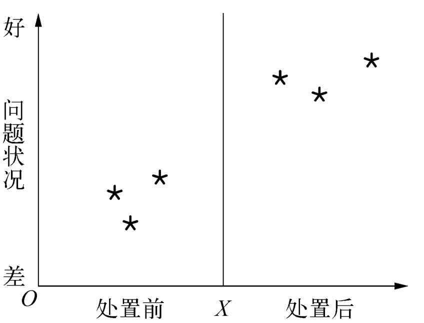


图 9-2 三次观察后的处遇结果

处置前的观察与处置后的观察当然可以不限一次。图 9-2 即表示处置前后的观察各有三次。坐落点虽不一致,但大体上与处置后的坐落点不同,因此更可以肯定处置的效果。

(二) $ABAB$ 型

此型的做法是先建立处置前的基准线(baseline)A(观察三次),第一条纵线

代表处置 X，处置后的第一阶段观察为 B（观察了四次），第二条纵线代表停止处置一段时期，再做观察，可以看到第二个 A（观察三次，其结果比 B 下降）；第三条纵线为第二次的处置，处置后的观察则为 B（共观察了四次，其结果甚至比第一阶段要略佳）。此型的设计又可称为抽回设计（withdrawal design）或反转设计（reversal design）（Herzog，1996）。如图 9-3 所示。

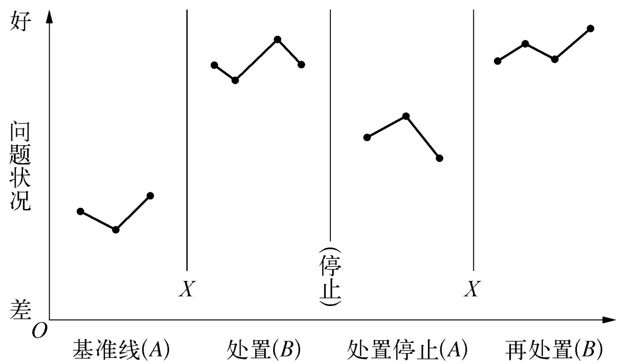


图 9-3　抽回设计示例

（三）多重基线设计

假设处置的目标不仅限于单一的标的行为，而为三种标的：上课迟到、在校打架、作业不交。纵线为处置的方法 X，但三条纵线都为同一种方法，如综合性的行为矫治法。X 左边观察值均为处置前的基准线（所以基准线共有三条），处置后的观察值则在右边，从图 9-4 中可以看出三种标的行为的变化。

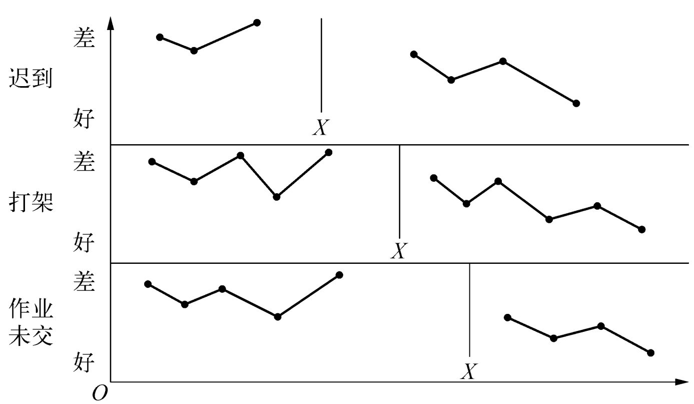


图 9-4　多重基线设计示例

（四）多重处置设计

若采用不同的处置方法，并且在不同的时间点依次进行，亦可以评估不同处置的成效，并可考量这些不同处置应以何次序出现（见图 9－5）。

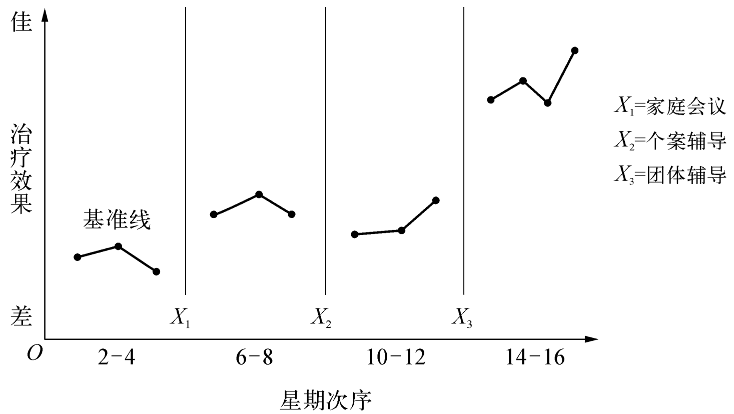


图 9－5　多重处置设计示例

四、单案研究的观察与解释

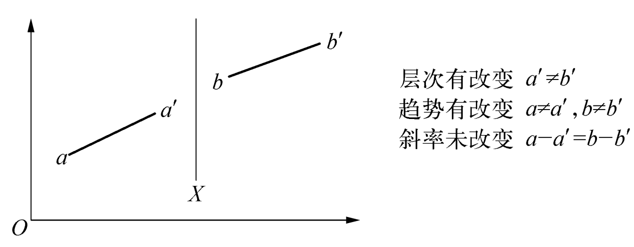
单案研究的成果经常借着简单的统计与图形来表现，可称为视觉的分析，一般都是看处置的层次(level)、趋势(trend)，以及斜率(drift)。

1. 层次(level)是指处置前后两个评估点的比较，如比较 $a' : b'$。

2. 趋势(trend)是指前后各自的方向性，借此来看改变的程度如何，如比较 $a : a'$的变化，或是比较 $b : b'$之间的变化。

3. 斜率(drift)是指改变的形态是往上或往下(较不考量改变的量有多少)，为方向的比较，如 $a-a' : b-b'$。

如图 9－6 所示。



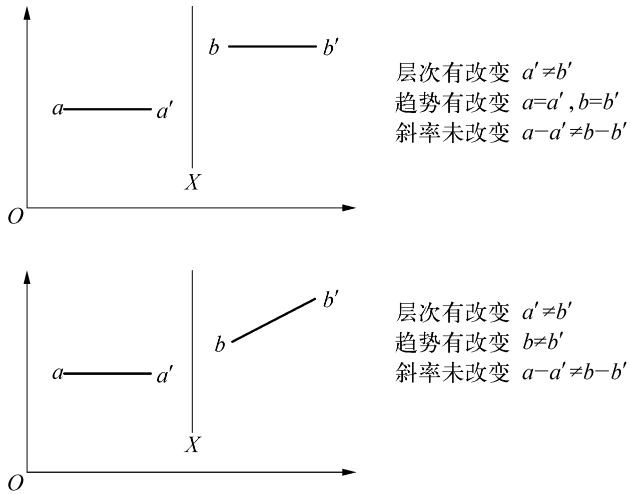


图 9－6　处置层次、趋势与斜率的视觉分析示例

五、单案研究的内在效度考量

若是用本章前面所介绍的 Campbell 及 Stanley 在实验设计中内在效度的条件来考量，单案研究所谓的处置的效应是否均由处置而来，能否克服历史、自我成长、测量以及情境不能控制等因素？答案可能不甚乐观。单案研究的科学性显然不如真实验研究严谨。单案研究因为只针对一个个案做研究，推论性当然低。然而，单案研究将是社工同仁“实务工作中不忘研究，研究中不忘实务工作”的绝佳伙伴，它不难，而且也清晰可行，最起码也可使实务工作同仁在工作中产生较理性的参考。

关键名词

前实验设计(pre-experimental design)：指实验“设计”无科学的严谨性，谈不上是个“实验”，只求方便或限于经费、人力进行的简单的研究而已。

真实验设计(true experimental design)：真实验设计的结构当然比前实验设计要严谨得多。最主要的是它拥有随机抽样，而且还有前测、后测，以及实验组和控制组的比较，它符合科学实验的要求，是真正的实验设计。

准实验设计(quasi-experimental design)：当某些情境不能用真实的实验方

法来控制变异量时,我们就利用准实验设计的方法,虽不能完全符合实验设计中的严格要求,但是大体已抓住了科学的精神与实验应有的态度,研究者必须了解有哪些特殊的变量是无法控制的,而对该控制不了的变量特别注意,以便使之更精确。

实验研究(experimental research): 是一种科学探究的方法,研究者操纵一个或多个自变量,并观察因变量随操纵自变量而发生的改变。

实验设计(analysis of covariance): 研究者为了探究某一问题而构思的一个扼要架构或策略,可借之简要说明抽样方法的变量安排、操作控制及进行程序,其内容通常以符号表示。

习　题

1. 何谓实验研究,实验研究又可分为哪两类,试分别叙述两者的优缺点。
2. 何谓前实验设计、真实验设计、准实验设计,其三者的差异是什么? 简述其适用的设计种类。
3. 请说明实验设计中随机化的过程及其在研究法上的目的。
4. 在实验设计中,对内在效度产生威胁的因素有哪些? 试论述之。
5. 一个好的实验设计,应考虑的因素是什么?

第十章

其他常见的研究法

第一节　前言
第二节　文化比较研究法
第三节　历史研究法
第四节　再次分析法
第五节　内容分析法
第六节　非干扰测量
第七节　结论

摘　要

研究本来就是探讨真理的过程，然而，探讨真理的方法应该是相当丰富与多样的，历史、文化、书面的资料，环境的种种现象，都可以给研究者在探讨真理的过程中提供很大的帮助，本章将介绍其他类型的研究方法，以供读者更广的参考。

针对某种主题，选两个或两个以上的文化或行为特质作为研究变量，以分析其相关性的研究方法，即为文化比较研究法，由于它不受限于单一的文化，所以在考虑的层面上也广了许多。

借着过去所遗留的事物或文献，以客观研究的态度去研究“过去”的过程，就是历史研究法。借着历史研究法，一方面我们可以学习社会工作专业过去如何面对与现今相似的问题；另一方面，也可以了解社会政策与社会问题之间的根源，使社工专业少走冤枉路。有关历史研究法的特质、过程，将于本章第三节探讨。

再次分析法是对现存的资料做更进一步的分析，在当今资源、时间有限的情况下，它是一种很好用的研究方法，然而，由于再次分析法主要是应用别人先前的资料所做的再分析，难免会有“捡便宜”之嫌，所以，如何进行再次分析，其目的、特质、优缺点，都值得我们小心谨慎。

只要有适当的主题与资料的话，内容分析法也是一种既节省又容易讨好人的研究方法，至于如何进行内容分析，需要符合哪些条件，将于本章第五节做介绍。

在调查访问的前或后，研究者在不干扰受访者，也不需受访者填问卷的状况下，借着对其他事物的观察或测量，使得到的资料可以佐证或增强调查研究时所得到的结论，此为非干扰性测量。其优缺点有哪些？读者应如何运用？本章第六节会加以介绍，提供给有兴趣的同学参考。

第一节　前　　言

把研究做得严谨是社工同仁应有的态度，问题是在现实社会中，如何能将人类环境控制得像实验室那样的严谨？尤其是牵涉到人类行为时，又如何把人规划得像物品一样，一切听命于研究者的摆布？当本质上不可能像实验室那样严谨，而研究者却也能把它“说”得如此严谨时，请问是科学？还是反科学？此外，人类的生存环境中，有一些内涵绝对不是实验或调查就可以寻出其因由的，我们必须用与实验截然不同的方法来处置，尽管答案或许不是那么的严谨与完美，却是合乎现实的，此即本章中所要探讨的文化比较法与历史研究法。这两种研究法是文化人类学家与历史学家常用的方法，并非社会工作者“分内”之事，但是这些方法的真谛，却可作为社会工作同仁们的参考。至少这两种方法都可以帮助我们对个案的背景与环境有更高的敏感度，在助人过程中可以提供更好的服务。

人类学家经常会以各地不同文化与背景的人或资料为研究的样本，借此来探讨为何不同的文化环境中，会有如此相异的行为或态度。如 Robert Sears 研究儿童断奶行为与情绪困扰之间的关系，他以 Kansas 市 80 名婴儿为例，发现断奶愈延后，情绪困扰愈大。这个发现很有意义，因为可供许多相关科系(社会工作、教育、心理等)参考。但是另一个学者 Whiting 却有不同的发现，他比较了 75 名不同文化的儿童，发现结果正好相反。在此我们不做谁是谁非的评估，但应了解有趣与复杂的文化比较值得社工同仁钻研。当今“社会工作本土化”在国内愈受注意之际，文化比较的研究更显出其特殊的功能。同学们从事研究时，常常在文献中可以看到很有兴趣的题目，虽然都是在国外的环境所做的研究，但若试图把此研究“搬回”国内来“印证”的话，这种文化比较的原则与方法更是我们可供参考的对象。

第二节　文化比较研究法

针对某种主题，选两个或两个以上的文化或行为特质作为研究变量，以分

析其相关性的研究方法即为文化比较研究。这种研究法有一些特质,有别于其他研究方法。

(一)是跨文化的研究

文化比较法不受限于单一的文化,它在不同文化中寻求一种类似的实验,以探索某种行为或态度的客观原因,此种研究法可以弥补其他研究方法的不足,在考虑的层面上"宏观"得多。

(二)与文化人类学关联颇深

文化比较研究法在抽样上牵涉到文化人类学的专业,进行此研究时,应细加研读其专业的抽样习性与特质(如民族志的使用),以免犯错,因为光凭随机抽样的知识,无法满足文化人类学对文化的定义与分类(就如随机抽样法中必须把抽样的样本单位定义清楚一样),否则若单位的定义都有问题时,再好的随机抽样也没有用。

(三)文化单位的考量

若要比较中美学龄儿童的问题处理能力差异,这个"中"与"美"的界定就令人煞费苦心。因为如何界定标准、相同的文化单位作为比较的标准可能有不同的定义与方法,John Whiting 建议以同一种文化内的地域社区作为比较单位,可能比复杂的文化比较有较好的效度。否则,在截然不同的文化背景中,硬要对某件行为或事物做比较,所得到的结果不仅争议性大,应用性也会受限。

(四)需考量因文化不同所造成的价值差距

对所观察、比较的事物如何量化与评分更是研究的重点,因为文化变量在不同社会中意义不同,分量自有差异。例如,东方传统文化的纳妾行为与穆斯林文化的纳妾不同,两者自不能用西方国家的尺度来衡量。

第三节 历史研究法

历史(history)从字面上而言,有三种不同的意义:(1)过去发生的事件;(2)历史是写下的东西,用来描述和解释过去的事实;(3)从过去所遗留下来的事物、所建构的报告。社会工作专业中所谓的历史研究指的是第三种,借着过

去所遗留的事物或文献，以客观研究的态度去研究“过去”的过程。

一、社会工作为何要探讨历史研究法

社会工作是助人的专业，为何要采用历史研究法呢？仔细思考后，我们会发现历史研究法其实对社会工作专业有莫大的帮助。从历史研究中，我们可以学习历史上如何面对与现今相似的问题。例如，借着研习宋代的救荒制度，我们可以了解积弱不振的宋朝如何做好救荒的工作，使得宋朝依然能够屹立三百多年，比穷兵黩武的秦朝还长得多。借着历史研究，我们也可以了解社会政策与社会问题之间的根源，今天我们在为各种福利法案伤透脑筋时，三百多年前的英国济贫法案（Poor Law）依然是很好的参考。社会工作专业做好历史研究，也可以增加有关组织如何成长和改变等方面的知识。例如，我们在为智力障碍儿童或轻度精神病患的安置犹疑不决时，研究社会福利的历史变革会发现，几百年来仍然跳不出以前早已存在的窠臼。有谓“前事不忘，后事之师”，做好历史研究可以使社工专业少走不少的冤枉路。

社会工作专业之所以也应对历史研究有所探索，主要原因如下：（1）社会工作专业也应对过去的事感兴趣，以便从过去的历史中得到教训；（2）社会工作专业应寻找现今的实务及其种种福利制度的根源，了解各种福利制度的缘起与背景；（3）社会工作专业研究者也应详加分析历史记载中种种的丰功伟绩或各样特殊的代表性行为，以提供给现代社会一个典范；（4）若一种福利制度有其成长与衰退，那探讨历史事件的兴衰原因也是社会工作专业责无旁贷的任务。

二、历史研究法的特质

社会工作专业固然要学习如何从事历史研究，但是所从事的历史研究法想必有别于历史学者所从事的历史研究法，基本上，社会工作专业所谓的历史研究法有其特殊的本质。

（一）它是批评性检定（critically examining）的过程

社会工作必须从研究的角度来分析过去所存在的记录。社工专业不从历史或文学的角度来探讨，他必须了解历史，而且不能用现今的眼光来看，要以当时的眼光来想象，这是对时间、空间、社会情境，以及人文制度、习俗的高度“同

理心”,为的是能客观地描述过去的事实。社工专业在面对种种资料与文献时,应比一般人更具备批评性的检定功夫。

(二)它具有非干扰性测量的研究态度

所有研究的内容是已经收集的资料,具有某种程度的历史性,因为它们以“过去”作为证据的基础。而过去是无法从直接观察中去测量的,因此历史研究的方法就需借着分析过去所留下的文件、物质及记忆进行。在这种过程中,为了能把历史过程与现今的状况做比较,研究者必须注意以非干扰性测量(unobtrusive measures)的原则来测量现今可测量、可比较的资料。因为非干扰性研究的优点在于受测者不知道自己被观察,而不会改变自己“正常”的反应,因而这种研究方法在历史研究中极其重要。

(三)它从过程中隐示其具有的意义

历史研究不仅是研究过去的东西,而是隐示一些“方法”与传承,我们也因此知道其过程演化中所具有的意义。以孔恩(Thomas Kuhn)的《科学革命的结构》(The Structure of Science Revolution)一书为例,孔恩在其生平中,最大的职业志向是研究科学的历史,他却也能从历史的研究中体会出“典范”的概念,指出科学的发展不完全是真理的延续与扩展,而是典范的原则在运作,科学革命必须在典范产生危机与不适应后才能产生。而这个论点,只有对历史的探讨极其深厚以后才有可能形成。

三、历史研究的过程(the process of history research)

虽然历史研究所研究的事件与现在有一段距离,不可能直接观察,但这距离的好处是我们能以较客观、较广阔的观点来看事件。它的研究过程可从下列三个过程来分析:(1)如何选定一个特殊可研究的主题;(2)如何评估所需的资料而提供客观的资料;(3)如何建构一个叙述来描述、解释发生的事实。

(一)第一阶段:选择研究问题

选择历史研究的主题时,最重要的考虑是个人的兴趣范围,当然也必须知道事件与事件间的一些联系。一般人在从事历史研究时,常常落入两个陷阱。(1)太强调“现在主义”(presentism):在选择研究题目时,仅仅以现在的重要性而论,认为过去发生的事件必须有现代意义才需做研究,其他与目前无相同性或相关性的重要信息与资料,均被排除在研究之外,这错误就称为“现在主义”;

(2) 古时主义(anti-quarianism)：这是另一个历史研究的极端，指只对过去的事件有兴趣，而不考虑其现实重要性及意义。仅把焦点放在使过去与现在不同的成分上，不管与社会工作专业的关系如何，不理研究过程中所需注意的理论性与实用性。

采用历史研究法在选择研究问题时，当然应该采取与社会工作有关的主题。有了这种主题以后，研究者再来找寻过去与现在之间的联系，如果是现在的资料较多，当然要以现在的资料为主、过去的资料为辅，对该主题进行比较研究。反过来，若是过去的资料多、现在的资料少，那就应该以过去的资料为主、现在的资料为辅，并设法提出如何将过去的资料及经验应用到现代社会。

(二) 第二阶段：收集资料

一旦选定研究问题，下一个任务就是要收集充足的资料来描述、解释过去所发生的事；我们的目的是要写一篇有关过去的报告，这篇报告要正确完整，除了能解释过去的事件外，也能对现在的实务有意义，为了达到这些目标，资料的收集就相当重要。虽然过去所留下的事件可能不完整，也可能都是片断，但我们要有技巧地选择特别有用的部分，处理这些不完全的资料，使之能达到研究的目标。

1. 资料的类型(types of source materials)

从事历史研究时，所能使用的资料范围颇广，其类别有以下几种。

(1) 出版的刊物。这些已出版的刊物能在图书馆中找到，也可借馆际合作获得，或在街坊书店中购买。

(2) 未出版的刊物。有些刊物并未出版，但被保存在特定的图书馆和档案保管处，也极有参考价值。对美国社会工作者而言，最重要的档案保管处是Social Welfare History Archives Center at the University of Minnesota, Minneapolis。

(3) 口头叙述的历史(oral history，也称为"口述史")。口头叙述的历史不只在老人的口中，若要探讨国内社会工作者的制度渊源，一些资深的社会工作者或社会福利主管也是重要的信息来源。

(4) 私人文件。某些对社会福利制度的推广极有贡献者的私人文件，像日记、未出版的报告等，若捐赠给档案保管处或图书馆，对未来的研究也极富意义。

(5) 社会福利机构出版和未出版的文件，如年报、简介、记录等。

(6) 报纸。报纸也提供不少社会政策发展的报道，包括社会机构的活动资料。若是研究者细心，或是方法富有创意，也会使研究有极重要的发现。

(7) 缩影片。有些报纸和口头访问的抄本制成的缩影片也颇有研究价值。

(8) 参考书目。参考书是我们开始找资料的很好的地方，借着著作的参考书目资料，研究者可以借之找寻更多其他的相关资料。

(9) 政府文件或听证会的报告，或如某委员会的研究、行政机构的报告、统计等。

2. 资料的选择

资料的来源固然不少，但使用这些资料时，并不是愈多愈好，所牵涉的是如何选择适当的资料。

第一，当研究者要使用一个文件当作历史研究的资料时，他必须先考虑事情发生时作者是否在场？假如作者当时并不在场，那他是怎么知道这件事的？若此事不能厘清，再多的资料也没有用。

第二，最好的历史研究是分析“原始资料”。原始资料提供事件目睹者或最初报告者的资讯；“次级资料”则是指分析原始资料后的资料，分析次级资料很少能再从过去得到新知识，而且最初作者的错误会永远存在。就如我们在准备研究的文献一样，若资料都只在别人的文献中寻求，而不能自己阅读原来的资料时，则资料一转再转，正确性会逐渐递减。

第三，评估所得到的资料。评估的方法如下：(1) 当时的作者有观察到他所报道的事吗？(2) 作者有客观的证据（诚实、不偏颇）和知识吗？(3) 文件是否在事件发生后很短的时间内完成，以至不会有记忆错误？(4) 文件所使用的字义是否不同于现在的字义？若以上这些问题都是“是”的话，则资料的可信度便会增加。

第四，考虑资料的内在效度与外在效度。文件内容是否一致？若所收集的资料相互矛盾，或在细节中相左，显然其内在效度太差。若文件中的资料和所知、所谈论的事件并不一致，或风马牛不及，则是其外在效度有问题。

3. 有关资料的一些特别问题

一旦研究的过程是在收集、分析“以往”的资料时，问题一定层出不穷。有

些资料可能是因为系统的选择，只阐述某一观点，因此容易产生系统的偏差。若使用的统计资料也有同样的问题，我们必须提防，否则资料一定是偏颇不全，推论时一定谬误丛生。

4. 资料不完整(fragmentary evidence)

资料不完整也是令人头痛的问题。历史研究法由于与当时现象的产生已隔一段时日，就算能收集到资料，也常会不甚完整，甚至与事实有所差距。研究若要“周全”实是相当困难，弥补的方法有限，在这种情况下，比较实际的方法是据实以报，坦白说出研究在资料收集上的限制。由于时空的间隔，历史研究法必须恪守“有几份证据就说多少话”的原则。此外，也由于历史研究法中所收集的资料以文献居多，而非研究者亲眼所见、亲身所经历，因此在推论时，研究者需要充分了解资料中所陈述的个体的心理、社会的文化，以及整个社会整体的考量，以便合理地贯穿这些推论与假设。除此之外，我们可使用社会科学理论作为这些差距的桥梁，把所观察到的事实与社会理论串连，不仅可以充实理论，也可以从理论当中找到与实务经验的差距。此外，在资料统计方面，因为所分析的资料本身不甚周全，或样本本身可能已有误差，研究法可利用统计技巧来估计可能的误差有多少。不然我们干脆就直接呈现不完整的资料，并说明它的缺点，来交代研究本身的不足。

(三) 第三阶段：确定资料的意义

在建构研究问题或假设时，通常需用某个理论架构或一组陈述来引导，毕竟用行为主义的观点来解释行为与用存在主义的观点来解释行为，一定有不同的论点与看法。因此一旦确定用什么理论来进行历史研究，该理论架构和假设一定要清楚，如此对于现象的解释才能有所根据，也才能决定我们要使用哪些证据，并确定这些证据的重要程度。

描述过去的历史事实时，一定要做整体的解释。而要了解事实的整体性时，我们必须先了解人类行为的命题，从这些命题中，研究者才可以更加明确资料本身所拥有的意义，才可以更有系统地解释人类行为。当然不同的理论会有不同的命题，不同的命题对行为的解释也有其差异。因此，研究的命题要清楚，告知读者是在什么命题下进行的历史研究，也因此，对何种现象就有何种的论点与解释，这种过程在历史研究法中颇为重要。

综合(synthesis)辩证的功夫对不同资料的处理亦有其功能。对符合命题

的资料固然要善加阐述,对于与命题相反的资料亦要加以珍惜,从正反两方面的资料中分析、综合,正确的资料才能出现。

(四)第四阶段:报告的撰写(writing the report)

历史研究报告的主要目标是借着研究方法的描述,来阐述过去的历史所代表的意义。因此,历史研究报告的组织必须详细解释研究的方法及其使用的过程,当然要对研究问题提出解答。历史研究报告要简洁,不要用太多术语,因为报告的目的在于沟通、讲解,而非考倒读者。绪论是历史研究报告中很重要的一部分,在绪论中,研究者应该介绍研究问题是什么、作者所使用的资料来源有哪些、可信度如何、研究假设和命题又如何等,使读者对整个研究的来龙去脉有所了解。研究报告的正文部分有不同的结构与叙述方法,然而,大部分历史研究的报告不是以年代为主,就是以主题为主,视研究的主题及本质而定。

四、结论

对很多社会工作研究者而言,历史研究法似乎距离甚远,感觉与社会工作没有直属关系,所以社会工作界中并不怎么关注它。然而,"前事不忘,后事之师"这个教训大家耳熟能详,探索历史可以增进人类的智慧,也可以使人避免重蹈不必要的覆辙。当社会工作专业在国内展开之际,当整个专业的发展借用了太多国外的资料之际,研究者在展望未来时,可能要花更多的时间与精力去研究过去。我国社会工作的历史发展的渊源在哪里?国外种种福利制度在现代化的历史发展中到底有何种使命与意义?过去的经验如何传承?国外的经验如何移植至我国?回答这些问题时,相信可以使我们更加了解历史研究的本质及重要性。

第四节 再次分析法

不可讳言地,研究有时真是大工程,从方案的研拟、经费的申请、文献的收集、研究问题与研究假设的修正、问卷的拟订,以及资料的收集分析、报告的撰写等,常使研究者呕心沥血,疲惫不堪。其实研究一定要严谨,但却不一定要弄得灰头土脸才算是仁至义尽。有时借着仔细规划,或借助现有的资料善加利用,也能做出一个很好的研究。再次分析法就是其中的一种。当社工实务或行

政的同仁因无时间、无人力去收集原始资料以供决策之参考时，若利用现有的资料或二手资料做再次的分析，则称为再次分析研究法（secondary analysis），这种方法若是做得好的话，不仅经济实惠，而且准确度上也不亚于花费颇多的实验研究。

一、何谓再次分析法

根据 Hakim（1982）的定义，所谓再次分析是“对某现存的资料做更进一步的分析，以呈现新的结论或解释的一种研究方法”。换句话说，再次分析法是一种研究方法，借由别人所收集的资料，把适合我们研究的资料再拿来做分析。资料的来源有大型的公立机构或各级县市政府，尤其是社会统计等机构的统计资料与文献等，其他一些公立、私立的小机构也会有重要的内容，如生命线及×老师的辅导统计，或一些残障福利机构对成员所做的调查等，也都值得我们根据其资料做再次的分析。

再次分析时，主要借由电脑技术的协助对现存资料进行另一层次的分析，如对案主的特性、活动，机构之成本、成果、服务输送及其他方面的种种可能变量（而这些变量却未在以前的研究被处理过），做较仔细或不同性质与方式的分析。简言之，用不同于过去报告的方式，对已存在的资料再做分析，以呈现的说明、解释、结论或新增的知识，即为再次分析法。再次分析法的运用始于 1970 年代末期，由 Smith 及 Glass 开始使用。社工领域中，Joe Fisher 曾于 1980 年代对社工领域的先前研究做再次分析，提出了几篇对社会工作实务界颇具震撼性的论文，引起了社会工作如何评估的论战，甚至也牵连了社会工作本质及哲学背景的重要议题，他的发现对社会工作的贡献不亚于任何其他研究方法（Hyman，1972；Kiecolt & Nathan，1985）。

二、再次分析法应注意的事项

由于再次分析法主要是应用别人先前的资料所做的再分析，难免会有“捡便宜”之嫌，所以在什么情况下才能做再次分析等问题值得引起我们注意。

1. 与原来的研究者有不同的重点和不同的研究问题时才可以进行再次研究，否则只是抄袭、利用别人的作品而已，此乃大忌，必须避免。

2. 针对相同的资料做分析时，应该用不同的方法与技术来分析，否则就变

成在查核或重复别人的资料而已。

3. 研究者应集中注意力,去思考有关理论性的目标或原先研究替代性的议题,甚至收集有关新资料,以做较仔细的比较,如此才可能对实务界及理论界有较具体的贡献。

4. 在进行再次分析之前最好与原作者联系,一方面征询其同意,另一方面也可以因此得到更好的协助,也可避免未来不必要的误会与争执。

三、再次分析法的分类

再次分析法若从不同的角度来看,也可以分为一些不同的类型。

(一) 从归纳和演绎过程来看

有的再次分析法较偏向归纳方法,有的则较以演绎的方式来分析。若是利用累积观察的过程去探讨研究的问题时,则可称为归纳式的再次分析法。例如,研究患精神分裂症的病人的复发率是不是与他家庭成员的情绪性反应有关,研究者便把调查对象的复发率与其家属是高度情绪性或低度情绪性(high levels of emotionality or low levels of emotionality)做相关性研究,因此便观察了很多个案,归纳出观察的结果,来确定两者之间的相关性,此为归纳式的再次分析法。用演绎法做再次分析时,以验证某个假设能否成立为主。如研究"大哥哥、大姐姐方案"(请大学生到中、小学与适应有问题的学生做朋友),若第一次研究的是案主(中、小学的适应问题学生)对志愿者(大哥哥、大姐姐)的态度会影响他们彼此关系的长短,而研究者根据这些资料做再次分析的研究时,则可用演绎法,去研究是否相同性别的志愿者对案主会产生社会心理功能的影响。此研究由上次研究的论点演绎而来,可称为演绎式的再次分析法。

(二) 从研究设计来看

从研究设计来看再次分析法时,有探索性、描述性及解释性三种。

1. 探索性再次分析研究的范围较大,有最多的机会去观察和解释。例如,可由现存的资料去检查不同变量之间的可能关系。这种过程不是去验证假设,而是从很多的自变量中去决定可能是哪一个自变量与因变量有关,或哪个自变量对因变量有较大的关联。例如,利用国际社会工作教育协会(CSWE, Council on Social Work Education)的资料去分析社工硕士人数减少的原因,发现情况最糟的学校是有博士班但师资欠缺,以及学校补助较多的学校,其社工

硕士班的学生反而减少。从这个探索性的再次分析中便得到一个很好的结论，建议学校降低传统硕士班的标准、减少资源的浪费，并提早做硕士班招募的工作。这种研究并不是真的在验证某种假设，只是从再次分析中，看是否能“收集”有关假设，以便更清楚探索事情的真相而已。

2. 描述性再次分析研究是用来描述同一时间内的单一变量，或设法把一个变量分成两个变量，但都只限于描述特性，而不是做“收集”假设或“验证”假设的工作。例如，CSWE 年刊中所刊载的统计资料，详细地列出国际社会工作界的教育状况，研究者若对这些资料有清楚的了解，则有助于其将这些资料做更翔实的说明，把一些资料串连做更清楚的解释，这种再次分析则为描述性再次分析法。

3. 解释性再次分析研究则以证明假设为重要任务，可从既存资料中对假设关系做简单的分析，也可设法控制某些以前的研究所没有控制的其他干预变量，来看看研究的假设及其变量与变量的关系会有什么样的变化，此为解释性再次分析（亦即借着再次分析的研究把事情的来龙去脉解释得比以前更清楚）。例如，从 CSWE 资料中，研究者把一些变量做验证的分析，看看学历、职位、责任，或不同的性别等变量会对薪水有何种不同的影响。如此就可以解释美国社会工作教育界中学者们的薪水受何种变量所影响。

四、再次分析法的研究内涵

（一）题目的选择（selection of a topic）

若要用再次分析法做研究，在题目的选择上就应该具有足够的企图心，最好是颇为特别的，或是范围较宽广的，否则很难引起别人的注意。当然研究者也应讲明为何必须用再次分析法，否则很难去说服别人。题目一旦选定，就必须马上考虑变量的问题如何操作，如何控制等。这些问题就是未来要招致批评的来源。

（二）寻找现存可用的资料（search for available data）

利用电脑寻找适当的学术机构或政府单位，因这两个来源的资料较完整，也比较容易在其他领域上发挥作用。当然一般人在做再次分析研究时，大体上都已经有了腹案，资料的来源也是十拿九稳时才会试图进行再次分析研究。否则，若先确定要做再次分析，然后再去寻找哪里有可以再次分析的资料，这样会

有本末倒置之嫌。

(三)重新创造资料(re-creation of data)

再次分析不是重复以前的步骤与资料,研究者首先必须应用旧有资料去创造新资料,也根据所欲采用的主题从旧资料中特别细心地寻找出所需要的资料。再仔细研究资料,若因细读、思考,能对某些变量的次数项目产生更深刻的了解时,研究的灵感便会如涌泉般出现。选择一些与研究主题有关的变量,但不能太多,否则会无法处理,因为一旦研究的灵感一来,常有一种企图,希望研究能回答所有相关的问题,所以就把所有的题目都拿来分析,结果弄得研究结果烦琐不堪。

(四)分析资料与比较结果(analyzing the data and comparing result)

再次分析是否做得好的主要功力就是它的资料分析。不仅要细心,而且还要有创造力与组织力,因为当务之急就是找出答案,让研究者与读者都能因为再次分析的新发现而对事实有更进一步的了解。

五、再次分析法所容易产生的问题

(一)信度问题

信度指测量的一致性。但某些资料与其测量本身就不一定是一致的。例如一个社工机构的督导被要求用一些标准去剔除一些“不适合”的案主,但社工每次在使用这些新的标准时,反应却有很大的起伏,每次的使用结果都不一致,可见这个“标准”的信度就有问题。或是机构内的社工同仁在使用这个标准时看法颇不一致,各有各的看法与认知,如此一来,信度也是不佳的。此时我们所谓的信度问题是指:当一个机构在使用再次分析时,以前所谓有信度的资料,在若干年后再拿来分析时,它的结果与以前的立论截然不同,如此一来便产生该标准信度方面的问题。

(二)效度问题

不一致(缺乏信度)的测量不可能会有效度,因为每次的测量都不一样,如何能测到所该测的?信度不足,效度当然就出问题。然而,在测量当中,有时即使有信度,也不一定有效度。测量工具的一致性是效度的必要条件,但非充分条件。在使用再次分析之前,要注意过去的哪些资料是我们现在仍然要用的,是否有把握能得到真实的资料。例如,研究金钱使用看法的问卷,二十年前物

质缺乏，即使效度再高，如今时代变得那么快，普遍使用金钱的态度当然大异于往昔，当初高效度的问卷，如今来看却不见得，这就是效度方面的问题。

（三）造成再次分析产生偏差的因素(sources of error)

为何再次分析会产生误差，除了以上所说的信度与效度问题以外，还有一些因素必须考虑。

1. 以前研究的工作人员若对研究不配合，排斥研究，而且也随便记录资料，这种资料再分析也没有用。

2. 不同的资料供应者对事件会有不同的定义与解释，若是原始的调查，我们可以马上查证，立即予以修正；但是再次分析时，先前提供矛盾信息的人，现在大多不知其踪，他们所做的不同解释或分类，再次分析的研究者也许会永远找不到头绪，当然因此会影响再次分析的准确性。

3. 资料供应者受个人喜好及社会价值观的影响，其资料会产生偏差，这些偏差在再次分析时丝毫没有弥补的余地。

4. 以前的资讯可能是合适的，但在再次分析时，可能定义上已有改变，分析起来，与第一次分析的定义可能不一致；以前的资讯可能是可信的，但现在对不同事情，可能也会有不同的解释，因而在何谓客观的解释时，在拿捏上颇不容易。不同机构或不同方案对同一名词、现象，采用不同的操作性定义，这样就无法做比较，这种研究的结果也可能与真实有很大的出入。例如，一般美国学校是大二分系，但是有些学校却是在大三、大四时才分系，因此在CSWE的统计人数中就会有出入，在师生比率上，有些学校就会有过多或过少的现象。

（四）遗漏的资料(missing data)

以前研究所遗漏的资料在再次分析时，一定会有更多的问题。在第一次研究时，研究者对遗漏的资料心知肚明，知道为何会如此，也知道如何去克服，如何去解释；但在再次分析时，研究者对这些遗漏的资料常会束手无策，在处理上一定很心虚，解释上也很脆弱，不容易克服资料遗漏的弊端。

六、结论

“再次分析”是在资源有限、时间有限的情况下很好用的一种研究方法，尤其当今电脑时代的来临，在收集资料的过程、储存、补救、分析等方面都很方便。

来日若有社会工作资料银行的设立,每个研究者愿意把当初所收集的资料、研究的过程、分析的代码与分析的过程,都录入光盘而存入社工资料银行供别人使用,也因此可以使用别人辛苦调查得来的资料时,那社会工作研究的空间必定更加开阔,再次分析法将成为一种相当重要的研究法。

第五节 内容分析法

为了对某问题做进一步的了解,所以针对其沟通或文件的内容,做较详尽的分析,则为内容分析法(content analysis)。试想,我们若把二三十年来的某报第一版面的广告当作研究对象,是否可从广告的内容、版面的大小、用字遣词的改变等项目一一来分析,看看是否可以因此看出时代的变化到底有哪些特质?广告方式的转变是否意味着传播文化的改变?诉求对象的变迁是否看出新生代生活格调的转化?这种从已有的资料与文献中所做的系统性及科学性分析,其实就是内容分析法。早期的内容分析法,依 Holsti(1969)的看法,需要符合四个条件。

一、内容分析法的四大条件

(一)资料的分析必须具备充分的客观性(objective)

在分析的过程中,每一个步骤都应有明确的规则与流程做依据。研究者应对所要研究的内容设计出客观的类别(categories),以便可以把资料“翻”成号码(coding),所以资料全部集中在一起时,就可以使研究者排除本身的主观性,而做客观的解释。这种过程,必须客观到若让另一个研究者根据同样的步骤来分析资料时也会产生与原来的研究者同样的结论方可。

(二)必须系统化(systematic)

研究者要把资料归类为某种类别或赋予其号码时,必须根据所规划的原则确定是纳入(inclusion)还是排除(exclusion),不能临时起意、擅自更改,必须完全依照既定的系统与原则行之。有了客观性与系统性后,也只能说具备了内容分析法的充分条件而已,还不是最标准的内容分析,它还需要第三个条件。

（三）必须通则化或定律化(generality)

内容分析所得到的结论最好与理论相关，或能形成某些通则，使它与其他理论有所关联。内容分析的目标绝对不是只在叙述某种现象而已，而是在分析某种因素与其他因素之间的关联；这些关联可能在文件中，可能在人们的叙述中，也可能在某些特定的人物与时段中。

（四）"量化的叙述"(quantitative description)

所谓"量化的叙述"不是指把所要观察的资料借着次数分配来说明某种特质，而是指用函数分析(contingency analysis)的概念，把某种现象的特质用数据的方式来说明其与另外一个变量之间的关系，或对其他现象所造成的影响。例如，Gilbert A. Williams 以电视广告为研究对象，专门计算电视节目人物及其对话中所涉及的"对暴力的倾向"与"对性的暗示"。首先，他用抽样的方法，从 *TY Guide* 杂志中选了 1980—1985 年的一些节目，然后他定义何谓"电视中的对话"(verbal referents)，及"暴力倾向"或"性暗示"所指为何，当他把节目样本一个一个计算其"性"与"暴力"的出现频率与次数时(当然，要归类节目内容是否合乎他的操作化定义时，有时必须用到质化研究的方法，以便能准确地把资料归类)，结果发现：性与暴力的确影响了节目的收视率与排行次序。同样的方法，我们也可以用来分析"最近几年来流行歌曲的歌词与青少年的心理适应"或"流行歌曲的歌词与某种行为特质之间的关系"；或从十年来的社会工作学术刊物作品中，来分析台湾地区社会工作的发展方向或潮流趋势；也可以从某个机构的文献资料中，来探索该机构的发展特质……这些都是内容分析法的应用。

二、内容分析法的内涵(Baker，1994)

（一）选择适当的主题

该主题必须是可以用内容分析的方法来研究的。当然该主题的分析事先也应有研究者本身的逻辑与假设，可让一般人了解：针对这些内容，用内容分析的方法去处理一定可以得到适当的资料，其结果与发现，也必将有学术价值与研究意义。

（二）从资料中抽样

当然抽样的先决条件是要使样本具有"代表性"(representativeness)，否则

一切分析都是空谈,所以随机的原则必须考虑。抽样的单位也要细加定义,否则会影响内容分析的品质与进展速度。

(三)从内容当中去发现资料的意义

内容分析法是要从资料中去探讨资料所携带的信息,那信息的来源必须考究,不能是“片面之词”。更重要地,在拟订题目以后,研究者应对所要研究的对象业已确定,所以,所谓研究中重要变量的研究者也应胸有成竹肯定资料的意义与价值。

(四)设计出“过录”(coding)的系统与原则

“过录”基本上就是把资料转化为数字的过程。例如性别有男有女,研究者可以把“男”定义为“1”,“女”定义为“2”。又如,为了探讨国内社会工作者的研究经验与社会工作专业未来的趋势,研究者用内容分析法收集了国内社会工作者十年来的研究报告,并逐一分析各个研究的特质,依类别来分,可以界定“儿童研究”为“1”,“青少年研究”为“2”,“中老年人研究”为“3”……“质化研究”为“1”,“量化研究”为“2”等,这些都是“过录”。当然“过录”的过程中,有时是以该资料的“次数(frequency)”“数量(amount)”“类型(typology)”“来源(origin of source)”“密度(degree of intensity)”等为主。若内容分析的主题是电视节目中的暴力渲染,可将“语言暴力”界定为是“1”,“肢体打耳光或身体推挤”是“2”,“斗殴”是“3”,“刀械攻击”是“4”……总而言之,就是要把所研究的内容系统性地转换为数字,而这些数字又能适切地呈现内容与意义。

(五)分析这些经过整理(过录)的资料

资料既然经过整理,就可以借此计算其次数、类别或其他特质,研究者便可以用表格或图形的方式,把资料的“意义”表现出来。此时,研究者可以发现,资料“过录”以后,在计算时便会显得较为灵巧。有时,可以把“过录”的原则更改,如把原来录入或编码的信息做另外的联结或增减,然后,将其进行结果分析后再与原来的资料做比较,也会有不同的发现与收获。

三、内容分析法的效度与信度

内容分析法固然可使研究者在耗费不多的状况下就可以做出不错的研究报告,但是对于内容分析法的信度与效度,研究者不能大意。为了增加内容分析法的效度,研究者必须特别注意所研究的内容是否“适合”做内容分析(并不

是所有的问题都可以用内容分析法去处理)，尤其要问什么问题，或看内容的哪些项目，都需要妥善考虑，因为问题与项目的设计牵涉到研究的假设与目标。内容或项目一旦确定，就必须好好规划“过录”系统，分哪些类别，计算什么特质……这些都是考虑“效度”时必须去注意的。至于信度方面，若是研究者把“过录”的标准与规则订好，使任何人看到同样的资料都可以有同样的归类时，其研究的信度或可增强；若研究的内容中有不少模棱两可的资料，使资料的归类或分析的方向产生不定感时，其研究的信度必然降低。

四、结论

只要有适当的主题与资料，内容分析法就是既节省又容易讨好人的研究方法。若要采用内容分析法，研究者应当对所要研究的主题有特殊的兴趣，加上手边又有不少的资料可应用，否则其内容就变成生涩不堪。虽然内容分析法都是以简单的次数分配方法在处理资料，但其精髓应是研究者的创造力与组织能力。有了创造力，研究者才能化腐朽为神奇；有了组织能力，内容分析法的真义才能有效地呈现出来。

第六节　非干扰测量

非干扰测量(unobtrusive measures)，“在调查访问的前或后，研究者在不干扰受访者也不需受访者填问卷表的状况下，借着对其他事物的观察或测量，使得到的资料可以佐证或增强调查研究时所得到的结论，此为非干扰测量”。(Webb, et al., 1966)

非干扰性测量源自 1966 年。Webb、Campbell、Schwartz 及 Sechrest 等社会心理学家在其共同的著作 *Unobtrusive Measures: Nonreactive Research in the Social Sciences* 一书中提出了非干扰测量的概念，认为非干扰测量可供调查研究时参考。

任何有研究调查经验的人都知道，访问受访者是使研究者颇觉疲累的一件事。研究者要考虑受访者的时间，必须设法排除与研究者本身时间上的冲突；要考虑受访者的文化程度，生怕受访者无法回答问卷上的问题；要担心受

访者的心情,就怕受访者一言不合掉头就走;还要注意访问时有否受到其他因素的干扰等。研究若能不影响别人,又可以得到必须得到的资料,何乐而不为?是以有这种“如意算盘”的研究,即为非干扰性研究。可惜,要特别注意的是,非干扰性测量不能取代访问调查研究,非干扰测量的结果只能供访问调查佐证或参考而已。

具体而言,非干扰测量有不可否认的优点:(1)不必干扰受访者,不必看受访者的脸色;(2)测量结果没有人为的伪装与不必要的影响;(3)不会形成实验研究中的受访者“成熟”“历史”或“测验”等影响研究内在效度的危机。

Webb 等认为非干扰测量可分为三种主要的形态:(1)物理线索(physical traces);(2)档案与公文(archives);(3)观察(observation)。但是截至目前,非干扰性研究的主要形态是物理线索的研究。在观察物理现象时,我们看其腐蚀状况或其物体上的黏着物,就可以知道该物所受的风化程度或所处的环境;在博物馆中观察一个区域的地毯磨损状况,就可以知道哪一个区域最受观众光顾;观察风景区垃圾桶内的垃圾数量,就可以分析出该区的消费行为(Rathje,1979);或是看一个区域的窗户(如福建沿海等地)就可以知道该区受海风侵袭的严重性;看城市市中心的空酒瓶就可以估计该市游民的酗酒状况;由图书馆内杂志的受损情形可得知哪些杂志有较高的“可看性”……其实,只要研究者心思敏锐(或是联想力不要太差),稍加观察生活周遭的种种状况,多少都可以找到某些问题的答案,这是非干扰性研究所要强调的要点。

非干扰性研究固然有其迷人的地方,但是考虑到研究必须具备的信度或效度,非干扰性研究法也有它无法克服的缺陷与困难。但是研究者必须了解,非干扰性研究不是用来替代某种研究方法的(严格地说,我们不应该把非干扰性研究当作研究方法的一种),若我们承认这种状况,而把非干扰性研究当作补充性资料来源而非主要的研究时(supplementary, not a primary source of data),那非干扰性研究对原来研究的帮助将会增大,因为它可以增加原来研究的信度与效度。从效度方面来说,研究有内在效度与外在效度,因为非干扰性研究所找到的证据可以使原来的研究更加有内在效度,同时也因相关证物的收集,因又能在别的地方找到同样的证据,所以原来研究的外在效度也因此可以大大提高。

第七节　结　　论

研究本来就是探讨真理的过程，显而易见地，探讨真理的方法不会只限于一种，只要研究者愿意，只要研究者不失去心灵的好奇，就能多方面去观察；只要研究者对本身拥有的资料深谙于心，就能对原有的资料赋予新的意义与活力；只要研究者不排斥过去，也不畏惧未来，那探讨真理的方法应该是相当丰富与多样的；历史、文化、书面的资料、环境的种种现象……都可以成为研究者探讨真理过程中很大的助力。

关键名词

内容分析法(content analysis)：为了对某个问题做进一步的了解，所以针对其沟通或文件的内容，做较详尽的分析，则为内容分析法。

文化比较研究(cross-cultural study)：针对某种主题，选两个或两个以上的文化或行为特质作为研究变量，以分析其相关性的研究方法即为文化比较研究。

再次分析法(secondary analysis)：对某现存已有的资料做更进一步的分析，以呈现新的结论或解释的一种研究方法即为再次分析法。

非干扰性测量(unobtrusive measures)：指在调查访问的前或后，研究者在不干扰受访者也不需受访者填问卷表的状况下，借着对其他事物的观察或测量，使得到的资料可以佐证或增强调查研究时所得到的结论。

历史研究法(historical research)：借着过去所遗留的事物或文献，以客观的态度去研究“过去”的过程。

习　题

1. 何谓再次分析法？再次分析法容易产生哪些问题？
2. 何谓非干扰性测量法？试述非干扰性测量法的优缺点是什么？
3. 试比较调查法、内容分析法、历史研究法等研究方法的优缺点。比较中应包括各方法的适用范围、资料特性、目的、研究取向与研究结果等。

第十一章

事后回溯研究与方案评估

第一节　事后回溯研究法
第二节　方案评估
第三节　德怀术(Delphi panel)研究法
第四节　行动研究

摘　要

在事情发生后，以调查研究的方式着手分析原因的研究方法即为事后回溯研究法(post hoc research)；如何进行事后回溯研究，其优缺点是什么，均将在本章中讨论。

社会工作研究主要以实务为主，绝大部分的研究都是事后回溯研究，所以事后回溯的方法，大家不能忽略。本章亦探讨了目前社会工作界以及企业界中最被广为讨论的主题——方案评估。社会工作为何需要评估？评估有何好处？评估可以绝对客观吗？如何按照事情的本质，选择最好的方式来评估？评估者应该扮演什么样的角色？在本章中将介绍需求评估、评估性评估、过程评估、成果评估以及成效评估等，以期社工实务中也能多多引用评估以提升实务的品质。

德怀术研究法以集合专家学者的意见为主。它借着匿名的方式，借着研究小组不断的问卷设计及专家们详细的问卷回复不断探讨，以求逐渐达成共识，使所拟订的方针能实质性地帮助机构未来的发展。

第一节　事后回溯研究法

一、源起

对过去所发生事物的研究，若是以文献事物为主时，为历史研究；若在事情发生后，以调查研究的方式，着手分析其原因的研究方法则为事后回溯研究（post hoc research）。因为有些过去所发生的问题或目前的问题，必须从过去寻求原因，而这种原因的寻求却不能用实验的方法马上找到答案时（如某个区域流行乌脚病），研究者只好分析既有的资料来探讨问题发生的可能原因、找寻其答案。事后回溯研究又称解释性观察研究（explanatory observational studies）或原因性比较研究（causal comparative research），因为需借着各种观察，提出各种解释（或先提出各种解释，然后从各种观察中找出原因），把种种原因与答案相互比较时，较合理的因素便会逐一出现。简言之，它是以回溯的方式，探究诸变量（原因）间可能的关系或效应的研究。

二、事后回溯研究的特质

事后回溯研究虽然是用既有资料来研究，但它仍然是具有系统性、实证性的研究法，因为它对所发生的事实提出一套系统的方案去观察、去解释，以期找寻可能的客观答案。事后回溯研究中所要探讨的自变量若不是既成的事实，便是根本无法操弄、无法控制自变量。回溯研究中，变量与变量之间的关系经常不能直接测知，必须从自变量与因变量之间的共同变异数来加以推论。严格而论，社会工作的许多研究都在事后回溯的范围内，因为社会工作经常是在某件问题业已产生后才进行研究。为了寻求解决之道，研究者往往在事情发生后去探讨问题发生的主因，这种过程就是事后回溯的研究。例如，1981 年前后，台湾地区青少年吸毒问题颇为严重，为了遏止这一势头，便要对吸毒青少年进行调查研究。研究时，我们必须先有一些论点或假设、前提，认为其吸毒是由青少年不知如何面对压力与挫折所致，因此研究的自变量就是青少年遇到的挫折或压力（这个变量是无法在实验室中重新再演绎或

设计的),因变量就是青少年的吸毒行为(这个变量是不能要求青少年再重复实施的)。为了探讨问题的答案,我们可以问那些吸毒的青少年当初吸毒的原因,特别强调是否当初遭遇了哪些挫折或压力。我们也可以以一般青少年为研究对象,一方面询问其平日所遭遇到的挫折有哪些,另一方面探讨其平日面对挫折是如何因应的(注意,这些青少年的挫折与吸毒青少年的挫折不尽相同;面对挫折的因应也不一定是吸毒行为,但是我们却努力去寻求这两个变量中是否有其共同变异数,而使我们能够据此来解释压力挫折与吸毒行为之间的关系)。

虽然事后回溯不是用调查研究的方法,但其却如历史研究法一般多方面采用当时的文献与资料,如统计记录、大众传媒及私人文件等。当然在收集这些资料前,研究者在心中已有其理论假设或问题的探索背景,研究者通常根据理论与假设去收集其所需要的资料,而不是临时有什么资料就根据那些资料来解释。事后回溯研究与历史研究之间的不同,是它仍然对过去的某种事件采取假设验证与访问调查的研究方法,而不是像历史研究法那样只根据各类文件去探索其前后因果而已。

三、事后回溯研究的步骤

事后回溯研究与其他调查的研究大同小异,只是有一些步骤较被强调。

(一)确认研究问题

就如其他的研究一样,它必须先确定研究的主题。在其他研究方法中,研究的主题可能是一种假设,而后用实验或调查的方法将研究的答案找出。事后回溯研究的研究主题则是以业已形成的问题为主,因此才有必要以事后回溯的方式探讨形成问题的原因。

(二)确立假设

有了研究主题,必须根据理论确立假设,事后回溯在假设与立论方面需要特别强调。因为问题既已产生,在推敲是什么因素造成这类问题时,不同的假设会导致不同的研究方向。在原因不明时,可能的因素林林总总,研究者却必须在推测的自变量项目中找出与因变量之间的关系,以证明研究的结果与假设有关,同时也必须在统计上得以证实。要注意的是,即使在统计上有了显著的印证,研究者也应小心谨慎,因为这个推测的自变量虽然业已证明,但仍可能只

是研究变量之间的共变数而已（意即还有其他变量与这个自变量一起作用，影响了我们所研究的因变量，只是这些其他变量并未或还未包含在我们的研究内）。

（三）选取比较组

借着理论或假设的指引，要探讨问题当中各种因素之间的关系时，即使研究变量得以证实，最好仍有一个比较组可以与实验组做比较，否则统计上的显著差异说不定只是偶然因素所造成的而已。事后回溯研究与实验研究一样会面临内在效度的问题，若没有比较组的对照，研究结果则较为堪忧。

（四）控制无关变量

就算有比较组可以对照，造成问题的可能会有其他一些无关变量，因此，如何控制研究进行时的无关变量，对研究者而言是很大的考验。通常较普遍采用的方法是配对法，它可使实验组与比较组之间的差异确定由实验的项目而来。比较同质组（与实验组性质相同的小组）或次级组（借着原来小组引申而来的小组），或借着共变数分析，来确定造成问题的各个自变量的功能，确定相互左右的共变数，也可以排除研究所不需要的杂质。

（五）资料的收集与分析

事后回溯研究的资料分析方法与其他研究方法并无太大的区别，其实分析方法是否妥当由什么性质的资料与我们要什么性质的答案而定。但一般而言，事后回溯分析研究较常用的分析方法分为以下几类：若是叙述统计则用平均数、标准差、皮尔逊积差相关、列联相关等；若为推论统计，则较常用 t 检定、ANOVA、卡方、Mann-Whitney U test、Wilcoxon signed-rank test 等。

（六）研究发现的诠释

事后回溯研究若能找出当初造成问题的一些原因，并且加以证实时，一定令人兴奋不已。但是事后回溯研究是问题发生以后，探讨其问题原因的合理化过程，就算颇具逻辑，也相当客观合理，但是研究者仍应抱着审慎的态度去解释，使整个事件的说明言之有理，对可能未探讨的、仍未找出答案的，以及在整个研究过程明显有瑕疵的，都能有清楚、诚实的交代。因为毕竟是事后回溯研究，所强调的是如何解释以前所发生的事，研究时难免已经有了先入为主的看法，要做到完全客观的解释并不是那么容易。

四、事后回溯研究法的优点

事后回溯研究法最大的优点是节省经费,问题毕竟已经产生,我们是做"合理的推测"而已,所以所用的方法、过程,所收集的资料均在我们的喜好、原始的情况与本来的背景中,不必做太大的付出,也不必在范围的拟订上煞费周折。第二个优点是可以看出时间系列上问题的长期发展,最起码可以比较现今与过去之间的差异,若是在研究设计中妥善规划,更可以因此看出问题在时间系列之间的改变,因此也是一种颇为有用的研究方法。

事后回溯研究法的严谨度似乎不如实验研究法,但也有比实验研究法优越的地方。

1. 变量使用的近便性。尤其当变量无法选择或操纵时,研究者必须直接研究某一问题的因果关系,但却不能选择、控制或操纵所需的变量,就可以使用事后回溯研究法。

2. 当所有受控制的变量流于不切实际或高度的人为化,以致妨碍诸变量间的正常交互作用,用实验研究法也无法找到确实客观的答案时,事后回溯研究法或许是个弥补的方法。

3. 当实验控制研究的目标不切实际、成本过高,或基于伦理道德的考虑,不应该也不能够使用实验研究法时,就可使用事后回溯研究法。

4. 事后回溯研究法也可获得关于某些现象性质的有用资料,如在什么条件下,什么因素以什么顺序与模式跟随什么因素发生,虽然这些答案并不是通过严格的实验而来,但是经过细致的事后回溯研究方法,把过程尽量设法予以控制时,所得的答案也是很有价值的答案。加上近年来,由于研究技术、统计方法以及部分控制设计的改进,使得这些研究的结果较具有防卫性。

五、事后回溯研究法的限制

当然,事后回溯研究法也有它的限制:第一,拥有资料者不一定合作,资料也不一定就确实,当今所用的工具与测量单位和以前所用的若是不同,可能会产生偏差,不符研究的要求;第二,它缺乏对自变量的控制,研究者为获得良好的结论,必须考虑一切可能影响结果的理由,或解释其他"似乎合理的对立假设",但如何证实其中的一种或多种理由较具有影响力却实非易事,而事后回溯

研究法在这方面更为有限，因为它对变量根本没有控制的能力；第三，事后回溯研究难以确定有关因素是否均已包含在进行的研究中，一切都是想当然的状况，无法证明是否与事实绝对毫无二致，这种结论当然不免偏颇，而且任何事情的发生，原因可能不止一个，而是由相当多的因素交互作用而成，且在某种情况下才会产生该结果，这种交叉作用对事实所造成的影响，事后回溯研究一点办法也没有；第四，导致现象的原因不一定必然是多元的，可能在某些情况由某个原因促成，而在另一个情况由另一个原因促成，此"单纯"所造成的"复杂性"，也必须考虑；第五，如能发现两个变量之间的关系，但要决定何者是因何者是果时常会有困难，而且，两个以上有关的因素，未必具有因果关系，它们可能与另外未被观察或认定的因素有关，若是事后回溯研究并未把这种可能列入研究规划，并尚未设法予以控制时，研究的准确性就会受到影响。

六、事后回溯研究法与实验研究法的比较

事后回溯研究法有别于实验研究法，从不同的角度分析，两者各有其优缺点，表 11－1 所示的比较可以看出其端倪。

表 11－1　事后回溯研究法与实验研究法的比较

	事后回溯研究法	实验研究法
自变量	无法操弄自变量	可操弄自变量
分派原则	因变量本身属性分类，由研究人员决定	可随机分派到不同实验情境中
推论	找出变量与变量之间的关联	可做变量间的因果推论

实验研究法的最大特质就是可以操弄自变量，确定了自变量、因变量，还可以控制自变量的数量，使一切变化都在操控中，因此可以断定到底自变量是如何影响因变量的。事后回溯法对变量的操作无能为力，因为当初造成问题的主因（自变量）早已消失，无法复返，更谈不上操控。实验研究法是以随机的方式，分配样本至不同的实验情境，以让样本可因不同的刺激有不同的反应，从而看出实验组与控制组之间的差距。事后回溯研究法则依研究对象本身的特质，由研究者决定分派的方式，其客观的程度自不如实验研究法严谨。实验法的研究结果可以做变量之间的因果推论，谁自、谁依、谁主、谁从，自有定论。事后回溯研究法顶多只能断定变量与变量之间的相关性，至于谁是因、谁是果，则无法断

定,事后回溯研究法所涉及的是关联,而非因果。

第二节 方案评估

一、前言

评估是目前社会工作界以及企业界中最被广为讨论的主题。社会工作为何需要评估?评估有何好处?评估可以绝对客观吗?如何按照事情的本质,选择最好的方式来评估?评估者应该扮演什么样的角色?这些问题都值得研究者注意。

二、评估研究的历史背景

20世纪以来,评估已成为一股风气,社会大众普遍认为,若要有效解决问题,社会工作机构需要有系统的评估。1917年时就有人提出,应该以实验的方法来研究问题、解决问题。但是一直到1940年中期,学术界大部分的“研究”都是在抽象的实验情境中进行的,极少是在实际的机构中实施。这种研究与社会工作的本质有很大的差距,所以这些(实验)研究的结果显然缺少客观性,很难应用在社会工作实务上。

设法运用研究的方法来解决社会问题,始于第二次世界大战后。当时社会结构巨变,家庭组合也受到影响,为解决城市的发展问题,因应因家庭结构变迁所引起的种种问题,大家逐渐明了——客观、务实的研究才是认清问题本质、有效解决问题的最好方法。20世纪50年代开始,已有研究者评估心理治疗对防止青少年犯罪的成效。20世纪60年代时,美国的期刊、研讨会,相关的专业协会等组织,也都广泛地以评估研究方法来检定多年来的做法是否客观与有效。20世纪60年代的评估研究着重于如何改善人类的成长状况,尤其着重于对早期儿童的教育方式与成效的评估,评估研究于是更加蓬勃发展。当时,社会政策制定人员、行政人员,以及社会计划者,开始认识到评估研究的用途,他们甚至认为评估研究可以应用到社会政策的发展与修正上(Weiss,1972;Riecken,1974)。

20世纪70年代开始,方案评估的研究空前鼎盛,而且社会大众普遍了解方案评估与做好政策计划、发展治疗方案等息息相关。当时很多研究经费都用到

有关贫穷方面的研究，儿童虐待、药物滥用的分析与防治，虞犯青少年，以及罪犯的防治方案的成果评估等。20 世纪 80 年代后，社会工作相关的机构更是处在一个处处讲求责信，时时讲求评估的时代。但是也有人认为评估的功能仍然有限，客观性仍然存疑，技巧也有待商榷，所以他们认为，与其把一大笔经费用在形而上的评估上，不如用在目标人口的直接服务上。因此，截至目前，赞成与否定两派仍然争论颇多。

几十年来，评估之所以愈被重视，其实有时代的背景因素。最主要的是经济因素的影响，当经济不景气时，所有的经费最好都花在刀刃上，每一分钱的使用最好都获得最大的效益，此时，对于方案的评估就势在必行，否则很多经费会变成浪费。此外，消费者运动也是促使评估研究日益普遍的原因。消费者（广义地说，应该是纳税者）认为有权利了解各种服务的来龙去脉，也有权利分析各种收费是否合理，评估研究也就显得愈加重要。而一个专业若既能考虑到专业的形象，又能顾虑到服务的品质，且还能分析其治疗的有效性时，方案评估当然会被更加重视，以便获得社会大众的信赖与支持。

三、评估的目的

之所以要评估，是因为评估可以给研究者许多方便。第一，它使执行者可以名正言顺地向外界宣布方案的有效性（effectiveness），因为方案是否有效业已通过评估。第二，评估也可以促使社工同仁提高工作效率，当任务进行时，若当事者知道他的工作会被评估，其压力固然会增加，但也会因此设法提高工作效率，以免比别人相形见绌。第三，评估也是展现新方案与技术的最好机会。若是一个新案要提出，在众人可能质疑甚或挑战的时候，最有力的方法就是提出新方案的成果评估，让质疑者马上可以看出新方案与旧方案之间的差异，若评估的过程严谨、方法客观，新的方案很快就可以被众人接纳。

我们也可以从研究的目的来区分研究的类型。从另外的角度分析，评估研究的目的不外下列五种：（1）基础性研究（basic research）；（2）应用性研究（applied research）；（3）总结性评估研究（summative evaluation）；（4）形成性评估研究（formative research）；（5）行动研究（action research）。这五种研究孰优孰劣，我们很难下定论，因为没有一种研究可以同时达到不同的研究目的、同时满足不同读者的需求，只有在目的清楚、读者明确之后，研究者才能做出具体的

研究设计，才能决定资料收集和分析的方法，以达成研究的目的。通常，基础性研究和应用性研究的“成品”以在学术刊物上发表的居多，因为他们的读者也是研究人员，均能以学术上的研究效度和理论的适用性等标准来判断这些研究的好坏。评估研究主要是向研究的捐助单位所做的报告，其将评估的结果作为决策的一些依据，以改进方案或是解决某种特殊问题（Patton，1995）。

（一）基础性研究的评估

基础性研究的评估是为了知识而追求知识的研究，旨在理解世界是如何运作的，他们对研究有兴趣是因为他们想知道这种现象的实质所在。他们致力于产生新的理论或检验现存的理论，对他们而言，理论是学科知识的概括或缩影。质化评估研究对于基础性研究的贡献，是通过立基理论（ground theory）（Glaser & Strauss，1967）来实现的，Glaser 及 Strauss 强调以归纳的方法产生和证实与经验世界有密切相关的理论。例如，为了有效研究台湾地区日益严重的外遇问题，研究者研究的目标是要形成造成外遇的基础理论，看看是否外遇者的人格特征、婚姻模式、问题解决模式有基本上的特质，这些以形成理论为基础的研究即为基础性研究。

（二）应用性研究的评估

此型研究目的在于帮助人们了解问题的实质，以增进掌握自己所处环境的知识。应用性研究的来源始于人们所遭遇的问题和所关注的事情，应用性研究者是在寻求基础性学科知识在现实生活中的实际应用方法，他们所评估的重点主要放在实务的应用性上。例如，在外遇的形成中，很多个案是因为不知如何处理在婚姻中所产生的问题所致，所以研究者就把研究重点放在个案的问题处理能力上，看看经过问题处理训练方案的个案是否有较高的婚姻稳定度，此即应用性的评估。

（三）总结性评估与形成性评估

总结性评估主要是对方案、政策的有效性或成果做一全面性的评断，以证明其理念本身是否确实有效，因而可以类推到其他状况中（例如，婚姻问题解决能力方案的施行结果到底如何？是成功还是失败？在方案进行结束后，要做一总结，此为总结性评估）。形成性评估则把焦点集中于特定的情境脉络中，其目的在于改进某一特定的方案、政策、人员或成果，其研究主旨是在“形成”某些方案，其研究对象与目的在于改进其研究情境之内的效果（例如，经过总结性评估

以后，研究者发现婚姻问题解决方案立意甚佳，但效果不良，因为当问题已冰冻三尺时，再让案主学习问题处理的方案效果有限，研究者认为预防性的问题解决方案要比事后补救的方式为佳，所以研究者把研究重点放在即将结婚的人，帮助他们做好预防性的婚姻问题解决方案，此即为形成性评估）。总结性评估所追寻的是在某特定范围或条件状况下的类推。形成性评估则无意寻求超越其研究之外的类推，其目的是对特定对象人员在其特定时间内所进行的特定活动进行评估，以寻求改进其行动效益的途径。形成性评估主要仰赖于过程研究、实施评估、个案研究，以及评鉴力评量，所以形成性评估经常需要质性研究做后盾，甚至于依赖质性研究来达成目标。

（四）行动研究的评估

行动研究是让方案或组织中的人员研究自己的问题，并设法去解决，因此，如何形成某种"行动"的研究目的，很明确地成为评估过程中很重要的一部分（Whyte，1989）。与形成性评估一样，行动研究主要是研究特定时间中的特定方案，其结论一般不超越其特定环境之外（例如，研究者为了鼓励更多的未婚青年参与问题解决方案，就举办了婚姻座谈会，邀请结婚多年且幸福美满的夫妻向这些即将结婚的人分享婚姻经验，此即行动研究，这类的评估即为行动研究评估）。

以上五种研究之间并无绝对明确的界限，但是必须了解不同的目的会产生不同的省察角度，产生不同的研究设计，并且会左右资料收集的途径，进而影响研究的成果，甚至连发表的格式亦有差异。所以把质性研究应用在评估时，妥协性便随之产生。无论是深度的研究问题，还是探讨广度的问题，两者之间如何权衡就是一个妥协的问题，这也是Cuba（1978）所谓的研究评鉴中的"边界问题"（boundary problem），所以在研究一开始时，确立研究重心与安排研究问题的次序，远比提出研究问题本身要困难得多。

四、方案评估的种类

依评估的性质与对象不同，方案评估有很多种类，主要有需求评估、评估性评估、过程评估、成果评估与成本评估等五种。

（一）需求评估（need assessment）

顾名思义，若研究的目标是以服务的需求量为重点时，此时的研究则为需

求评估研究。在需求评估研究中,研究者可以以问题的严重性为前提,讲出某种方案的重要性;也可以以案主本身的需求为范围,借着案主的需求量,来确定方案的重要性。若以问题的严重性为前提时,研究者必须界定问题如何存在于方案中,而且指出问题的严重性,告知读者马上进行其方案的必要性,以便减少受害范围,增加受惠人口。若是以案主的需求为着眼点,则应分析出案主的需求项目、需求量、需求程度等,最好也借着统计的分析精确地指出何种案主需要何种方案,如此一来就可以有一正确的需求评估,在提供服务时,也能使有不同需求的案主都得到满足。

进行需求评估时,也经常会遇到一些问题。第一种常见的问题是评估不准确。评估得太高时,所造成的浪费可想而知;但是万一评估得太低又容易失去需求评估的意义,等于需求评估失败;需求评估一不准确,服务的有效性一定会大打折扣。根据经验,需求评估最大的困难是它无法考虑成本效益。若问案主需求什么,他们当然会有很多需求,问题是这些需求是主事者可能提供的吗?知道问题的严重性,借之以估计解决该问题所需付出的代价,这种程序非常合理,但是如果问题太严重,解决问题的代价也着实太高时,那这种需求评估也只是画饼充饥而已。

造成需求评估不很可靠的原因之一,是研究者对问题的发生率估计过高,预估的问题的严重性也比事实夸张,如此需求评估当然不准。第二种原因是事先的个案记录没做好,致使在进行研究时无法有一个较为客观合理的基本底线(base line)资料可供参考,所以研究计划的量度或预测当然也因此整个产生偏差。第三个原因是缺乏标准化的换算或估计原则,致使对案主的估计不准。例如要进行不幸少女(雏妓)的需求评估时,研究者必须估计某地区内有多少不幸少女,但是要确定什么是不幸少女时,研究者可能要大费周折,因为官方的定位与私人机构的界定差距太大,甚或可以说是南辕北辙,如此一来,需求评估如何进行?所以在进行需求评估研究时,可能要有周全的事前准备,计算需求的"标准"是否业已准备妥当?如何界定案主?如何计算其需求单位?更重要的是,支持研究的单位,未来根据需求评估所做的方案或治疗的预算到底有多少?若是预算只有一点点,那需求的调查层级就不要太高,否则这种需求评估根本一点也不实际。预备进行需求评估时,事先一定要把与该问题最直接、最熟识的重要人物(key person)找出来,好好访问他、请教他,并且请他就研究者所预想

的问题做估计，坦白说出其意见，是否有太不切实际的现象产生，以设法提高研究的可行性。

（二）评估性评估（evaluative assessment）

事实上并不是所有的事件都可以做有系统的评估，要做评估之前最好先进行“评估性的评估”，先界定问题的定义清不清楚？机构本身的概念合不合理？对问题的认知客不客观？所要进行的研究本质能否量化？能否适切地评估？评估性评估强调在评估进行之前先要评估机构目标是否清楚、客观，机构是否有不适当的概念化过程等。

很明显地，若是某些方案的定义无法具体被操作；或是方案本身没有特殊性；研究目标不清楚，也显现不出来进行评估的必要性；或是处置时间的长短没有标准化时……研究就不能通过评估性评估的考验，当然进行评估时一定会遇到问题。

因此，若要进行评估研究，机构管理者与评估研究者必须保持密切的联系，共同把研究的目标、概念、各种操作化的定义等界定清楚，在方案设计阶段，研究者也应该与社工同仁多多沟通讨论，看看是否有实务存在，而研究者可能还一无所知的问题。机构若要进行评估性评估，整个服务的流程应该相当清楚地被界定，若流程图无法画出，最少也应有实务的阶段或周期可供参考，一旦流程清楚，每个员工知道如何进行服务，研究者也因此更可根据此流程来制定评估的标准，而后的种种评估工作也才得以进行。

把机构运作的流程界定清楚，固然是做评估性评估的先决条件，另外一个方法则是用目标管理法来处理。所谓目标管理法（managed by objectives，MBO）就是由宗旨（goal）演化为具体可行的目标（objectives），再由目标化为活动（activities），活动执行完毕时，其成果亦可测量（outcome measures），也因此可以断定该机构是否可以做评估性评估。所以当机构的整体宗旨清楚，该宗旨又能衍化为一些具体可行的目标，再设计为一些活动，最后又可以加以测量目标的达成率有多高时，该机构的方案就可以进行评估性评估。

例如，以“青少年观护所”内的青少年为例，该“观护所”设置的宗旨就是使偏差行为的青少年借着“观护所”的种种辅导方案可以恢复正常，进而早日离开“观护所”。我们为这个宗旨设计了三个目标（objectives）：第一，减少该青少年群体本身危机的再出现；第二，提高该青少年群体的自尊；第三，减少该青少年

群体的侵略性行为(aggressive behaviors)。针对第一个目标,活动的设计可以为:提高该青少年群体对问题严重性的认知(如告知其相关的法律常识),教导问题解决的技巧,设计户外活动使其精力可以发泄等。针对第二个提高青少年群体的自尊的目标,其活动可为:使青少年参与有关社交技能训练的团体,设计有意义的休闲活动。针对第三个目标,其活动可以设计为个别的咨询及约谈等。当然在这些活动进行时,我们一定要安排一些方案督导及负责的社会工作者等。最后则是成果测量。针对第一个目标,我们可以以青少年顺利结业安全返回原生家庭的时间长短来决定成功与否;针对第二个目标,则以某专门测量当事者自尊指数的量表来测量;至于第三个目标,也有类似的 Achenbach child behavior checklist 来计算其侵略性行为是在何种程度。如果一个机构的方案可以用上述条例来演练,显然地,它必定可以通过评估性评估。

进行评估性评估时的问题不少,以下一些措施则可多少解决其问题。例如,在行政体系中,有人专门设计各个成员的角色及各人的工作内容(job description)。审核其机构政策时,要看其是否有可能获得资源?是否有其他团体的支持与合作?机构的持续能力又如何来评估其政策是否妥善?若要评估机构本身的运作状况,则看它的发展潜力、管理能力以及考核方法的可行程度。不管如何,机构本身应有研发人员专门研究“未来”的机构发展,也一定有人专门负责处理“当今”服务输送的进行。机构的考核评估者应与方案执行者区分开来,以减少彼此之间角色的冲突,毕竟机构愈大,事务愈多,不可能某个人是多方面的专家,分工合作仍是必然趋势。

(三)过程分析评估(process analysis)

从评估研究的角度来看,过程评估与成果评估一样重要。进行成果评估时,研究者假设若方案执行前后之间有明显的差异,这种差异必然是由方案所造成的改变所致。但是,除非研究者清楚知道方案是如何进行的,方案的内容与特质又如何,否则他不能斩钉截铁地就说前后的差异必然是由该方案所造成的。研究者在进行结果(方案的有效性)评估之前,必须对过程(方案的执行过程)评估有所了解。

所谓过程评估,是“检查并测量自变量进行的每个步骤与细节是如何造成相关因素的变化的”。为了对自变量做有效的测量,整个方案的进行过程必须操作化、标准化,对于方案影响所及的种种现象与情况也必须设法予以测量,如

此，才可以确定此时所谓的"过程"是什么，过程的功能又如何。

"定义"不难，"执行"却不易。真实的状况中，要把治疗的过程详细确定，而且又把治疗的功能仔细评估常要大费周折。因为治疗的"一致性"非常难，针对不同的个体，要使治疗标准化根本不切实际，等到治疗有了结果，又说该结果是从何种过程而来，也是不太能令人信服的，因此，过程评估的难度可想而知。实务经验中，要做好过程评估，最好由评估者自己直接观察，这样比较容易了解每个过程进展的结果。评估者也应该在评估的进行中随时检查各种报告，以便了解方案的施行所产生的种种变化。

评估者若能随时收集当事者本身的工作记录，并与整体的测量做比较，过程评估的准确性一定可以提高。了解被评估者每天的工作内容，如与案主接触的次数、地点、活动的类型、经费、时间等资料，有利于评估时的查证，因为研究者因此可以掌握每个过程的进展，也可以减少评估的盲点。

其实质性研究是探讨过程评估的有效方法。找一些个案(包括治疗有效的和治疗无效的)，问他们在治疗进行时是何种感受，是何种认知，是如何得到帮助的，当然也包括他们本身是如何抗拒、如何更正、如何改变或放弃的。当社会工作者了解案主在整个被治疗的过程中的心路历程或是一举一动时，社会工作者一定可以得到有关"过程"的种种信息，这种评估必然有很高的说服性。

(四) 成果分析评估

方案评估中最常见的就是成果评估，事情一做完，到底好还是不好？值得或是不值得？成果有多少？都应该好好检定，这就是成果评估。从研究法的角度来看，好的成果评估要克服三个问题。

1. 适切的研究设计或研究规划。既然讲成果，就是要看实行方案以后的成果到底有没有比未实行方案的成果为佳。但是成果是否一定由那个方案而来，一定需要进行非常周全的研究设计后才能确定。好的研究设计会让方案的功能达到最高(实验变量所造成的变异数愈大愈好)；会控制其他非属实验范围的变量(控制实验外的变量)；会减少误差变异数，使研究与调查的误差降到最低。要同时达到这些目标就需规划良好的方案设计，其中的随机抽样原则更是不可忽略。

2. 好的成果分析评估必须具备准确的量表，能将方案前与方案后在各方面的成果都准确地测量出来。

3. 分析研究的成果若是丰硕,它的前因后果及其以后所可能带来的影响等若都能分析叙述,则可使受评估者知其所以然,也可因评估对其未来的服务更有助益。

成果评估的研究设计最理想的方式是实验设计,因为实验设计极为严谨。但是实验设计的随机原则却会使研究者相当辛苦,主要是因为受访者不是物品,把受访者“随机”编往实验组或控制组,在人情上、伦理上都说不过去,但若不遵守随机的原则,研究的客观性又值得怀疑,因此研究者须斟酌实情、仔细判断、小心为之。所以成果评估在规划研究设计时若能尽量避免有“不治疗”的控制组出现,或可减少这种困境。成果评估在设计量表时也要妥善为之,因此有关增进量表有效性的方法必须好好掌握。

1. 确定所选择的工具是否与机构特定的目标匹配。选择测量工具时,不应只考虑准确度,其制作过程的哲学或文化背景亦不可忽略。

2. 当测量工具很弱或适用性不高时,不妨借着多种资料来源与方法配合测量。

3. 规划成果评估时,尽量避免用“再发率”作为评估的标准,因为若是计算案主的再发率,就必须考虑在哪个时期以前进行测量,更需考虑当事者返家的那段时间有何干扰因素(非实验因素),如此一来,研究将会没完没了。

总之,成果评估的可用性、普及性均广,值得应用。但是成果分析在报告中必须有系统的整理、系统的解说,避免一言以蔽之的草率与笼统。成果评估更应说明研究进行的背景与限制,如样本的大小,在测量时所面临的种种困难,机构的特质及其在受评估时所可能造成的差异等都应在评估中说明,如此才能有科学、客观、合理、合情的分析。

(五) 成效分析(cost-benefit analysis)

看方案是否有效果是成果评估,但是为了达到方案的效果到底付出了多少代价,就牵涉到成效分析的问题。在研究中,考量一个方案所得到的效益与其为此效益所付出的成本之间的比重即为成效分析。在经济压力下,如何提高效能固然重要,但是考虑所付出的成本更是刻不容缓。一个好的方案不仅能按照计划得到所预期的效果,在时间与金钱上的付出也不能超出底线太多。否则就算成果评估良好,成效分析却会太差。

社会工作(或是一般社会科学领域里)一涉及成效分析时就问题不少。因

为社会工作界中的很多处置和所得到的效益很难衡量，到底付出了多少代价也很难被计算。例如，一个试图自杀的个案，经过三个硕士学位的资深社工每个星期各四个小时的辅导，最后却仍自杀身亡。请问这种个案到底值不值得处理？三个资深社工如此为个案奔波，是否值得？若整个办公室八个社会工作者一个月中服务了一百二十件情绪困扰的个案，而只有十个个案的情况有所改善，是否这家辅导中心应该马上“关门大吉”？这些问题处处都牵涉到成果评估与成效分析的困难与盲点。

成本效益分析所谓的“贵”或“便宜”，“值得”或“不值得”很难定论，最简单的方法就是做相对性或比较性的成本效益分析。例如，要做甲机构（某地区的生命线协会）的辅导效益分析时，最好先找出一个性质和规模都与甲机构相当的另一个机构做比较，他们彼此之间的地缘、人口、规模都相当，所聘用的社会工作者人数也相同，如此我们就可以对这两个机构的该年预算、所辅导的个案数、个案对该机构的满意程度等做比较，看看他们所服务的个案效果是不是与他们所付出的经济成本成比例。如此做相对性的比较，就可断定哪个机构较为精简、效益较佳，哪个机构在服务的“量”方面可能有不足的现象。

不管相对性的成效分析如何进行，社会工作者必须了解，服务个案的时间成本非常难以计算，是否值得也很难定论。社会工作同仁必须更加注意，服务方案的间接成本或社会成本也很难计算。就如某个机构的募款活动，花了一些人力募了多少钱固然很容易计算出来（活动的成本效益），但是谁又能计算募款以后，公司形象会变好或变坏（变好或变坏又值多少钱）？这次募款成绩斐然固然令人高兴，但会不会因此产生一种疲乏的作用？会不会让以后的募款活动愈做范围愈小？而这些变动就是它的间接成本，至于某种在媒体大为表彰的行为、游行示威、某个专业人士行为不检等，会造成专业社会成本的付出，谁能计算出是多少经济代价？

研究者在进行一个机构的效益评估时，就算能找到某个类似的机构做比较性的效益评估，也应该注意各个机构的单位成本不一，无法在同一的角度来比较。纵使机构的地缘、规模、社工人数等都相仿，但是可能甲机构的理事长有较丰富的经济资源，所以成本上他可以不必那么认真地去考虑（就算表面看起来贵了不少，但是付出的成本却可以在理事长的另一个事业里变成收入），请问这种成本我们如何计算？由此我们可以知道效益分析的难处。

效益分析固然有其困难,但也可用一些方法去克服。例如,设法把一些待评估的事项先“标准化”。针对某些案主做成本效益分析,研究者对相似的案主最好用相同的标准做比较。区域医院在多少病床数时应该有多少医务社会工作者?每人每周应该处理多少个案?若用这些标准来计算一些医院,或许可以找出一个常模或基数,也可以当成标准化的参考资料。在固定的一段时间内,以预算的角度计算每个个案的成本与效益,而不要瞻前顾后地考虑太多因素(若长期评估每次都得到修正时),在一段时间以后也可以找到适用的标准,这也是效益评估可以参考的方法。

五、评估研究的步骤

评估研究也有其步骤。在分析这些步骤之前,读者应该了解,有的作者认为评估研究是范围非常广阔的研究法,可涵盖实验及准实验的方法(Babbie,1992)。但在本书中,评估研究只是众多研究方法之中的一种而已。其步骤如下所述。

(一)形成研究问题(formulating the problem)

研究问题的形成与其他研究方法大同小异。评估性研究则大体以某业已进行一段时间的个案为对象。为观察个案的效果,方便决定日后是否继续进行,评估性研究须据此形成“问题叙述”部分。

(二)测量(measurement)

1. 确定所要测量的结果是什么(specifying outcomes)

若要对某医院的医疗社会工作的服务效果做评估,首先必须先确定评估如何去测量?要测量什么?是病人对医疗的满意度呢还是遵守医嘱的程度?是住院天数的缩短还是医疗成本、费用的降低?一旦测量的重点有所不同,整个测量的对策也会有所差异。

2. 测量实验的整个背景(measurement experimental context)

测量不只是测量因变量的变化状况,更应考虑到实验的内在及外在背景,因为这些背景因素也会影响因变量。例如,评估婚前辅导的效果时,不能光测量受访者婚后的适应状况,然后找一个控制组来比较;可能的话也应考虑到那个时期的社会经济情况,若是景气不佳,普遍就业情况不良,新婚的人即使接受过几个小时的婚前辅导,其效果可能还不足以“对抗”大环境的压力。这种大格

局的考虑，在评估研究中也必须注意。

3. 把干预（处遇）定义清楚（specifying intervention）

评估研究中若是把样本分成实验组与控制组两种，研究者对实验组施予干预后，将其结果和控制组的结果做比较，这种方式似乎很简单，其实不然。就算实验组优于控制组，是什么因素造成这种差异更需研究者明察。说不定不是干预的效果，而是干预过程中的某种人为因素。若参加婚前辅导的人在结婚以后两年内的夫妻适应显然优于控制组，研究者想当然地认为一定是干预的结果所造成的影响，这样未免太武断，因为说不定只是施予婚前辅导的那个"人"在辅导期间用了很多宗教上的训诲与说辞所致，而不是婚前辅导的内容。所以，评估研究中，研究者有义务要把干预过程的内涵、过程详细解释，才能做好客观的评估。

4. 把其他变量清楚交代（specifying other variables）

在评估研究中，研究者主观上当然要把干预的策略定义清楚，其他可能造成样本改变的因素也必须做好分析与说明。这些变量的掌握一方面可由样本的属性变量如性别、年龄、教育程度等来探讨，另一个方法就是对不能从主要干预策略得到改变的样本进行好好分析，看看是否有什么特别的因素造成施行处遇的障碍。这两种方法都可以找到造成改变的主要因素以外的其他因素。

（三）实验设计

评估研究也可以界定为实验研究。但是实验研究可以分成很多类型，在本书中，引用了很多 Campbell 及 Stanley（1973）的观点，两人把研究设计分为真实实验设计（true experimental design）、准实验设计（quasi-experiment）两大类，各类又可分为一些不同的类别与设计，本单元不再详述。

（四）详细分析为何成功，为何失败

评估研究的主要目标固然是对方案的执行做个总结，确定方案是要继续还是中止。但是站在学术的观点，更重要的是找出成功或失败的确实原因。测量一个方案成功与否其实还不算难，把为何成功、为何失败的原因清楚地列出才是真正的功夫。若再考虑成本效益的因素，评估方案成败的定义更可能见仁见智——从人性的观点或伦理的角度来看是一种结论，从经济的角度来看可能就截然不同。但是若能客观地分析成败的因由，总会对方案未来的发展有帮助。

六、评估者的角色

方案评估者的角色不应当被界定为一个“计算者”,而应该是一个“咨询者”,计算者无时无刻不拿着算盘,斤斤计较地计算一切支出与收获,难免会变成一个只有数字而没有人性的冷血人物。咨询者却是一个懂得机构状况的行政行家,也是一个社会工作的专业人士,他知道行政上科层体制的特质、优点与缺点;他也知道社会工作同仁的爱心或无奈;他更能体谅改变一个体制所要付出的种种代价。咨询者熟悉基层,因为他从基层来;他更知道数字,对数字敏感,却更能通晓数字的特质与意义。基本上能对机构或社会工作者做成效分析的人,应该至少是个资深的社会工作督导,否则他只能拿着数字讲一些与实务毫无关系,一点也不能对机构有帮助的批评而已。

评估者难免会有角色冲突的现象,其原因如下。(1) 每个人的观点与价值系统都会有巨大的差异,没有一个人在做方案评估时可以很肯定地说他的看法是现实唯一的解释。若是评估者没有这样的认知与心态,倒霉的一定是受评估者,会有不知所从的困境。因此,方案评估者应该有足够的沟通技巧,了解同事们的看法,也让别人分享他的所见所闻。评估者在立足点上应当站稳,他知道所为为何,因此他应该有说服别人的能力,更必须有接纳别人看法的雅量。(2) 讨论机构工作的质与量孰轻孰重时,千万不要有众人必须一致的企图与动机,一牵涉到机构的质量问题,研究者彼此都会有不同看法,彼此多了解是必要的,改变别人则大可不必。(3) 评估一完成,面对受评估者的失望情绪,或受评估者对评估结果不可置信、相当不服气时,方案评估者则要有相当的雅量去接纳,了解什么是必须坚持的,也知道什么是无奈的;评估者更必须接纳自己,否则评估者的角色一定非常尴尬。

面对评估过程中经常遇见的种种问题,评估者应当有一些基本的认知:他必须了解,评估不是可操生杀大权的利器,评估研究与实务之间的很多事物其实是相互影响的。规划(planning)、管理(management)与评估(evaluation)三者其实是密不可分的,评估者不应当高估自己,在专业面前,不同的步骤与方法各有各的重要性,且密不可分,评估者并非高人一等。为了减少评估者与受评估者之间不必要的紧张,机构进行评估研究时,最好能请机构外并且熟悉机构运作的人来担当较佳;一个了解机构的外人(outsider)在执行评估时所面临的

抗拒应该较少。成熟的评估者会让自己在评估过程中尽量扮演学习者(learner)或是老师(teacher)的角色。说是学习者,是因为在评估别人的过程中也可以发现受评者的优点,评估者可以从中学习。说是老师,是因为可能本身的辈分较高(与受评估者之间以前也可能是师生关系),被机构聘请来执行评估,在评估的过程中,他不是扮演裁判者,也不是操生杀大权的执行者,而是一边评估、一边教导的老师。裁判者是"把受评估者的错抓出来",老师则是"把受评估者不会的教给他",两者在意境及气氛上差距甚大。若能扮演老师的角色,可让评估的过程多几分温馨。当然,评估者也应随时提醒自己,在整个评估过程中,一定要扮演"客观观察者"的角色,否则,他就失去了评估者的身份。

评估者和实务工作者之间常常发生关系的冲突,主要源自彼此的角色欠缺澄清,尤其是评估者本身与机构的负责人或创办人之间的关系,也应该厘清。例如,评估者是受创办人之托来进行评估,那评估者是不是也应六亲不认地对机构负责人进行评估?或是他应该与负责人事先就沟通清楚,他只是进行对日后方案推广有所帮助的辅导与评估而已。这些暧昧的角色若不事先言明,在评估结果出来以后,一定会使彼此的关系既紧张又难堪,谁也搞不清楚彼此之间应该怎么相处。此外,评估的结果与发现如何发表也应事先言明。有人认为,有关评估资料方面,资料拥有者(受评者)与评估者之间都有对外公布其评估研究"发现"(findings)的权利;有的则以为只有机构负责人才是有权力发布结果的唯一人选,评估者受聘于他人,不应该径自发布研究结果,何况,评估结果会影响很多当事者的生计与升迁机会,不可不慎。

七、结论

评估经常会影响机构的政策决定、经费配置,以及该机构的专业责任度,评估结果当然也会对机构的主管造成威胁。评估由机构的负责人担当或委托外来者评估(in house evaluator or external evaluator),如学术或研究单位来处理,所造成的影响或困扰都有所不同。国内中央或省级的社会福利单位对下属单位经常要做评估,以决定经费的补助或得奖人员的选拔,有时是内审,有时则委托外界学术单位来处理,经常则是以组合式的委员会来担当,因为如此可以免除不必要的人情压力。但是仍然要考虑到是否能保持绝对客观?评估者是否对事项相当了解?评估者本身与受评估者之间有无一些评估以外的关系存

在(如师生关系、亲戚关系或利害关系等),否则评估的信度与效度仍然会受到影响。团体压力也是影响评估的重大变量,尤其是国内的人文与社会特质,在这方面很难清楚了断。若又牵涉升迁或经费的分配时,理性与非理性的抗争都会增多,因此进行评估时不可不注意。

第三节　德怀术(Delphi panel)研究法

一、前言

Delphi(特尔斐,或译“德怀”)是古希腊的某地名,是神明向人晓谕神旨的所在地。语意上来说,Delphi 的原意是 Oracular of Delphi,意即深奥难解的、含义不明的、意义模棱两可的含混不清的话(mysterious or unclear because more than one meaning is possible),panel 则指专家小组。最早使用 Delphi 方法是在 1950—1960 年代,为了弥补量化研究科学性的不足,研究成员 Olaf Helmer、Norman Dalkey 与 Nicholas Rescher 对研究方案(Project RAND)进行了不断的修正与重新建构(modification & reformulations),建立了 Imen-Delphi 的研究步骤。

所以所谓的 Delphi panel 很清楚就是把一群专家小组集合在一起,他们各有不同的意见,他们的语意也常让一般人无法彻底了解,有时他们彼此各不相让,他们的时间也常兜不拢(在某些议题中,他们宁愿匿名,而不愿公开表明是他个人的意见),因此把各个专家的意见整合在一起,以便有效地解决问题,这就是 Delphi panel 的起缘。

二、Delphi panel 研究法的特质

(一)定义

Delphi panel 研究法是集合专家意见的集体决策技术,针对机构内所面临的问题,集合专家的知识和智慧,借着不断的意见整合而后设计问卷调查,进而使这些专家们可以达到共识的研究方法。这是以集体的方式处理复杂事务,为评估现状、策划未来、提升政策品质,或做业务转型诊断的研究方法(Linstone &

Turoff，1975；谢卧龙，1997；吴雅玲，2001；游家政，1996）。

（二）特质

Delphi panel 研究法与一般的研究方法或社会调查法不同，它有如下的特质。

1. 匿名性

若把专家学者集合到一起，有时意见反而不容易整合，因为囿于人情或避免冲突，或专家学者会迟到早退，致使会议无交集、无结果。而 Delphi panel 是用邮寄问卷的方法，专家间对争议的问题表示意见时也无具名，如此一来，每个人就都可以畅所欲言，不受人情、时间、人数多寡的限制，这是 Delphi panel 研究法的特质之一。

2. 须有适当的问卷设计

为了整合专家小组的意见，工作小组（机构内的承办同仁）首先就必须把问题摊开，问专家的意见为何。他们的意见若有不同的地方，工作小组收到回函后，就必须马上整合再设计第二次的问卷发给各专家，如此不断地进行下去，直到意见全部整合为止。可见，若无问题整合及问卷设计的能力，此研究法则很难进行。

3. 以书面形式表达意见

借着回复工作小组所设计的问卷，每位专家的个别意见表达都应是书面形式。针对不同的意见，也应用书面表达自己的看法，如此一来，参与的专家都有充裕的时间思考问题，所表达的意见当然也有较佳的品质。

4. 是相互回馈的过程

从第一次的意见表达开始，专家们的工作就是回复问卷所列的问题，而这些问题就是工作小组整合各专家们不同意见后所设计出来的，专家们不仅可以表达，还可以不断地再度思考，他们也必须面对别的专家所抛出来的问题，如此一来一往，直到意见趋于一致为止。这种回馈的特质应是 Delphi panel 最大的特色。

5. 是不断反复进行的步骤

工作小组采用多次连续问卷，以逐渐降低彼此意见的差异，直至专家的意见可以被机构采用或是可以有效解决问题为止。理论上来说，专家群的意见总比个别专家的意见要好；经过挑战依然能够屹立的意见总比坚持己见但仍然一

意孤行的意见为佳;经过反复不断的流程所得到的意见总比一时激动的意见保险。

6. 专家取向

专家的意见是主髓,所以如何选取有代表性且具有专业经验的专家非常重要,最好这些专家还具有在该专业内的不同背景,如此一来,才可以收到开始时的“百花齐放”,最后则是“万众一心”的效果。

三、Delphi panel 研究法的步骤

(一)成立工作小组

主办单位本身必须先成立工作小组,确定研究的目标,还必须详细规划专家小组应包括哪些人选、如何与他们联络等。问卷返还时,工作小组更必须整合意见,设计问卷,借着问卷设计引导专家表达对机构的实质意见,这是重要而且不可或缺的工作。

(二)选取参与研究的专家

Delphi panel 的研究经常都需时两三个月以上,每次的问卷都必须由这些专家仔细研究、详细回复后,才能投入使用,所以专家群的选取相当重要。专家们最好能拥有相关领域的实务经验及技能,一定要有参与的热诚与兴趣,最好人格上也喜欢与人分享意见,回复问卷则须有文字沟通或参与别人意见及阅读数据的能力,否则无法达到研究的效果(谢卧龙,1997;游家政,1994; Volk,1993; Delbecq, 1975)。

(三)设计问卷、整合问卷

问卷一般是开放性的,使专家能自由发挥、随意表达。关于第一次的问卷,必要时可以根据文献资料来设计问卷。专家回复后,工作小组则须整合问卷,简单统计所得到的各专家的意见;工作小组把意见整合后,让专家们做第二次或第三次、第四次的回复。为了让专家们对别人的意见有所回复,工作小组不妨把专家的不同意见列出,然后再让专家们对这些不同意见依其“重要性”来评比或排序,以便使专家们的意见逐步可以达到共识。

(四)量的研究及统计的研判

借着计算专家们对各种方案、不同意见的统计,如平均数、评价等级的和(等级和愈低,重要性愈高)、各方案的平均等级及意见的离散程度,专家们可以

知道其他专家的想法，也可以在问卷中表达折中后的想法。

（五）质性分析

借开放性的问卷了解每位专家对事情的看法（质性研究中已介绍其方法），将所有资料诠释、分析、归纳与整合。

（六）结果分析

Delphi panel 研究当然是最后面的结果最具代表性。在最后的问卷中，专家们所表达的意见可以说是研究的成果，因为这些意见已经过多次整合，已融合各专家们彼此的看法与意见，当然最具重要性。

四、优点与限制

从实务的观点来看，Delphi panel 研究法非常实用，因为每个参与的专家都能独立自主地表达自己的意见，研究的过程也是专家们仔细参考别人意见后理性表达的过程，结果更是每位专家的整合，这是解决机构问题或是拟订机构未来发展方针很有效的研究方法。这个研究法颇具经济性，但为确保专家的问卷回答维持相当的品质，最好事先就让专家们知道他们应有的待遇、责任、义务与权利（给付专家的费用千万不要太寒酸，否则专家们不认为他们有义务要帮你做作业，但其实专家小组人数有限，总比调查上百上千的个案所花的费用要节省）。

不能否认地，Delphi panel 研究法也有不少限制与缺点。首先，专家的选择不是那么简单，合适的专家本来就不多，考虑到他们的参与意愿时必然更少，整个过程耗时不短，不是写了一次开放性的问卷就大功告成，两三个月间专家的流失问题也必须考虑。此外，在处理分歧意见时，不仅专家本身会有坚持己见的现象产生，连工作小组如何整合问题成为下次的问卷也会各有各的看法，至于理论基础及架构上的薄弱，就更需要未来的研究者不断去钻研与规划了。

第四节　行动研究

行动研究（action research）是让方案或研究团队的人员在执行方案的过程中一方面执行方案，一方面又能使执行方案过程中产生的问题再进行研究并解

决问题,行动研究的目的显然成为方案改变历程的一部分(Whyte, 1989)。影响所及,研究和行动之间的区别变得相当模糊,研究方法也不那么系统和正式,但研究与方案的组织成员与问题之间则更为接近(Patton, 1995)。

行动研究的研究设计并不是很正式,资料的收集也较为因时因地制宜,研究涉及的人员有时也直接参与资料的收集,并进而自行研究。行动研究的成果常在内部就地利用,以解决方案、组织或社区内的特定问题。行动研究的范围也经常只是该组织或社区内的特定问题,而不是一般研究所强调的研究推论性或严格的抽样过程。

简而言之,行动研究的目的是解决方案、组织或社区中的问题;行动研究的重心是组织与社区的问题;预期结果是立即行动,尽快解决问题;行动研究的类推程度仅是此时此地;行动研究的主要假定是处于该场合的人能借由研究解决自身问题;当研究结果发表时,它发挥作用的形式是研究参与者之间的人际互动,是非正式的,此外,其结果甚至是可以不发表的;行动研究好坏的判断标准是研究参与者对研究和问题解决历程的感受的真实性与接受性的高低,以及研究结果的可行性高低。

四五十年来,台湾地区的社区发展加上社区总体营造多少已显现出社区居民意识的抬头。一方面,社区居民对社区本身的问题愈加敏感,一方面,他们对社区生活品质与格调愈加重视。面对未来,为了提升社区生活品质,也为了维护社区居民自治的原则,行动研究是很贴切、很实际的研究方法。

关键名词

成效分析(cost-benefit analysis):在研究中,考量一个方案所得到的效益与其为此效益所付出的成本之间的比重即为成效分析。

事后回溯研究法(post hoc research):在事情发生后,以调查研究的方式着手分析原因的研究方法称为事后回溯研究法。

评估性评估(evaluative assessment):评估性评估强调在评估进行之前先要评估机构目标是否清楚、客观,机构的运作是否有不适当的概念化过程等。

过程评估(process analysis):即检查并测量自变量进行的每个步骤与细节是如何造成相关因素的变化的。

需求评估(need assessment):研究目标以服务的需求量为重点的研究即为需求评估。

习 题

1. 何谓事后回溯研究法？并说明其在研究上的优点。
2. 试述事后回溯研究法与实验研究法的异同。
3. 试述方案评估研究的意义及评估的标准,并举例说明何谓客观性低而主观性高的评估。
4. 社会工作为什么要进行评估研究,其优缺点是什么?
5. 何谓需求评估？何谓评估性评估？试举例说明之。
6. 方案评估的种类有哪些？试分别简要说明其意义与适用时机。
7. 若您目前身为实务评估工作者,可能会面临哪些难题,该如何克服?

第十二章 抽 样

第一节 前言
第二节 抽样的概念与术语
第三节 如何进行随机抽样
第四节 抽样
第五节 结论

摘　要

抽样是研究法中相当重要的一环，若研究的抽样工作有瑕疵，整个研究的价值便会大打折扣。因此，在社会科学的研究中，研究者如何进行抽样是相当重要的。

决定抽样好坏的最重要的因素就是抽样是否在随机的情况下进行，因此，随机的理论与抽样分配原则十分重要。一般而言，抽样可以分为随机抽样与非随机抽样两种。随机抽样又可分为简单随机抽样、系统随机抽样、分层随机抽样、整群随机抽样四种。

非随机抽样又可分为具备样本抽样、配额抽样、立意抽样、滚雪球抽样、便利抽样五种。问题是，在什么状况中应该使用何种样本，又该抽取多少样本才能使抽样真正代表总体……这并不是可以简单回复的问题。本章则用一些例题及经过简化的数学运算，使读者能更清楚了解抽样的操作及其在统计上的意义。

第一节　前　　言

所谓“抽样”是指在研究问题确定后，研究者决定要观察谁、观察什么，以及决定由谁来回答问题的过程。具体而言，所谓抽样，第一，抽样是从总体中抽取一部分而已，而不是把总体中的每一个样本都拿来调查。总体可以是人，如调查一群学生、家长或是案主等；样本也可以是某件事，如虐待小孩的事件、夫妻吵架的频率，或是性生活满足的程度等；样本也可以以某种组织或机构为主，如调查某机构的人事状况、收入与职业之间的相关性研究，或是案主对机构的满意程度等。第二，抽样必须能代表总体，否则就不是好的抽样，所以分析抽样的样本以后，若是取样的结果是真，总体也一定是真，在样本里的结果一定可以在总体当中存在。第三，如果取样是随机的（而且抽样的样本不是寥寥可数），则可以不必担心样本当中有太特殊的情况，因为借着大的样本的抽样，我们有充分的理由相信，这些太特殊的样本是可以互相中和的。

抽样最大的好处是经济实用。1984 年，美国盖洛普（Gallup）公司于美国总统大选之前进行了调查，该公司只抽选两千名样本为调查样本，就可准确地预测九千万投票人的投票行为。就是因为盖洛普的抽样做得很好，所抽选的样本相当具有代表性，整个调查过程也都非常严谨，使调查有其客观性与正确性，由此带来的调查经费节省更不在话下。若非如此，不经抽样而要调查全美投票者的话，那就必须花费太多时间及金钱，而且，调查的结果不见得比只调查样本的结果来得准确（人多事繁，加上旷时废日，不免状况百出）。所以，为了得到一个正确性高、省钱又省力的研究，把抽样做好便是首要任务。

第二节　抽样的概念与术语

一、概念与术语

一般而言，抽样有三大关键要注意。一是样本结构问题，样本本身有哪些

重要的内涵及文化因素,是同质或是异质,如何散布……这些问题都与样本的结构有关,抽样之前,此结构因素必须考量。二是抽样的过程,如何把样本中的个体组合起来,使其既有经济性又有代表性,这是抽样中所必须考量的。三是样本设计的细节,即找到客观、合适、合理的样本后如何使误差尽可能减小也是需考量的关键。因此,抽样时不妨先厘清一些重要的概念与术语。

(一)元素(element)

研究资料收集与分析的基本单位,如虞犯少年(以该少年为调查的"单位")、社团(以该社团为单位,如调查其组织架构)、学校、种族(以学校或种族为分析的单位)等都可作为元素来进行研究。不同研究所做的抽样其元素不同。

(二)总体(population)

系固定研究单位与元素的集合体,如以虞犯少年为元素时,其总体是某地区内所有的虞犯少年。

(三)研究母体(study population)

样本实际被选取的架构中所有元素的集合即为研究母体。例如,研究者若研究虞犯少年,总体则为全省所有的虞犯少年(若为 3000 名),但抽样时,我们的抽样设计只拟调查其五分之一,为 600 名,但只有 520 名完成调查,所以此 520 名则为研究母体。

(四)抽样单位(sampling unit)

不同抽样有不同的抽样单位。如在某市做选举行为调查时,开始研究者在八个街道中随机取四个街道,此时是以街道为抽样单位。在所抽选的四个街道中,研究者又在各街道中随机抽取五个居委,再从每个居委中随机取两个社区,然后在各社区中抽选十户为调查的样本。在这不同的抽样过程中,我们有不同的抽样单位,分别为街道、居委、社区、户等。

(五)抽样架构(sampling frame)

根据研究主题与对象的范围,列出一份包含所有合乎抽样条件的名单,再从这全部的名单中,决定抽取百分之多少,此为抽样架构。

(六)观察或资料收集单位(observation unit)

根据抽样的单位与架构,研究者就有了观察的单位(如观察虞犯少年的行为或幼童的侵略行为等),但是进行资料收集时,有的研究可以由别的成员来代替,例如,我们访问虞犯少年的管理者,或问幼童在学校的老师与在家中的家

长，如此一来，观察单位与资料收集单位就会不同。一般而言，资料收集单位与分析单位通常是相同的，如研究者可能以从家长那儿（收集单位）收集的幼儿资料为研究分析的重点，所以此时的收集单位就是分析单位。

（七）变量（variable）

依不同数值或类别出现或变动的一种属性（property），如性别（类别变动代表不同的性别）、年龄和收入（依数值的变动代表不同的年龄与收入）。

（八）统计数（statistic）

描述样本性质的量数则为统计数，如样本群的平均受教育程度为十一年，大约为高中二年级，此十一年则为有关样本的年龄统计数。

（九）抽样误差（sampling error）

总体的实际值与样本统计数估计值之间的误差为抽样误差。最好的抽样当然是百分之百代表总体，但是现实状况中抽样的结果经常会与总体略有出入，此为抽样误差。

（十）信赖程度与信赖区间（confidence levels and confidence intervals）

估计总体的实际值时所制定的误差判定标准为信赖程度（alpha 值）。而信赖区间则是总体的实际值在某一概率估计范围内的数值分布区间。

二、总体的界定（defining a population）

在取样之前须对总体做清楚的界定，才能决定以什么样的抽样才可代表总体，而为了界定总体，最好的方法是能先确定定义所要用的指标，如果指标不明确，那么什么是总体也无法详细厘清，而总体不明确时，完美的抽样便无法进行。若是贸然进行抽样，则在推论中会造成“生态性的错误”（ecological fallacy），即把某一单位的分析结果作为另一单位的结论（硬把从分析样本中所得到的结论认为是总体的特质），准确性当然令人怀疑。一用到抽样，最值得注意的就是它的代表性（representativeness）。在研究的过程中，会有一些我们关心或感兴趣的情形及现象，我们可以称之为“变量”（例如，研究对儿童的虐待行为时，父亲或母亲自己童年是否也曾遭遇过虐待是个重要的变量）；变量一定有两个以上的分类（level），如有、无被虐待，被虐待的现象也可以有不同的类别，如身体虐待、性虐待或疏忽等。如果“取样”能代表总体，则样本在各变量上的变化或分布情形就可视为总体的情形。事实上，样本能完全代表

总体是很难的,最重要的方法是在取样的过程中增进随机的机会,并且把抽样的架构做得周延、完整,则样本与总体之间的差距便会逐渐减小,其代表性也必然提高。

第三节　如何进行随机抽样

一、从一个饼干的例子谈起

假设你在饼干店工作,店里卖三个品种的饼干,有巧克力、花生及椰子,是以重量(千克)来计算价钱的。这15份饼干的重量如表12-1所示。

表12-1　十五份饼干重量(千克)

巧克力		花　生		椰　子	
饼干号码	重　量	饼干号码	重　量	饼干号码	重　量
1	2.5	7	5	12	5.5
2	3	8	2.5	13	4
3	2	9	4	14	6
4	3.5	10	4.5	15	3.5
5	4	11	3		
6	2.5				
合计	17.5		19		19
平均	2.92		3.8		4.75
总重量	55.5				
总平均	3.7				

二、统计的运算

在表12-1中,我们可以得到一些统计数字。

(一)算数平均数(arithmetic average)

1. 椰子饼干的平均重量最重,巧克力饼干的平均重量最轻。

2. 15份饼干中,最重的那份饼干是6千克,最轻的是2千克。

3. 花生饼干的平均重量比其他饼干更具代表性。因为巧克力饼干的平均重量是 2.92 千克，椰子饼干的平均重量是 4.75 千克，花生饼干的平均重量是 3.8千克，而总平均重量为 3.7 千克。

（二）变异数(variance)与标准差(standard deviation)

变异数的计算原则是计算每个测量值与平均值之间的差异。在饼干的例子中，可计算每份饼干与平均重量的差异，将这些差异值平方后并相加，再除以饼干数即可。变异数的平方根即为标准差。以表 12－1 为例，其变异数的计算如下：

$$\begin{aligned}&[(2.5-3.7)^2+(3-3.7)^2+(2-3.7)^2+(3.5-3.7)^2+(4-3.7)^2\\&\quad+(2.5-3.7)^2+(5-3.7)^2+(2.5-3.7)^2+(4-3.7)^2\\&\quad+(4.5-3.7)^2+(3-3.7)^2+(5.5-3.7)^2+(4-3.7)^2\\&\quad+(6-3.7)^2+(3.5-3.7)^2]/15\\&=19.4/15\approx 1.29\end{aligned}$$

所以标准差则为$\sqrt{1.29}\approx 1.14$

（三）饼干样本的抽选

若无法把全部 15 份饼干都拿来称，研究者就必须用抽样的方式来估量全部 15 份饼干的重量，假设抽 5 份饼干，并由此来决定全体的平均重量。此时，研究者可将 15 份饼干制成卡片，并打乱这些卡片，然后随机抽选出 5 张卡片，此即为样本。

（四）抽样误差(sampling error)

上述抽选出的 5 个样本其平均重量与 15 份饼干的平均重量会有差异，此差异即为“抽样误差”。而“非抽样误差”则是在非抽样过程中产生的错误，如因秤本身有问题等产生的误差都可称为非抽样误差。抽选的样本愈大，则抽样误差愈小；然而样本过大的研究反而会产生较多的非抽样误差。

（五）从相同总体中抽选重复样本

我们可从 15 份饼干中重复抽选出 5 份饼干，从 15 份饼干中可以抽选出很多种 5 份饼干的组合，而每一种组合皆可计算出一个平均值。假设我们从 15 份饼干中抽取了 100 种 5 份饼干的组合，并绘成图，横轴代表 5 份饼干的平均重量，纵轴代表相同重量样本的次数。见图 12－1。

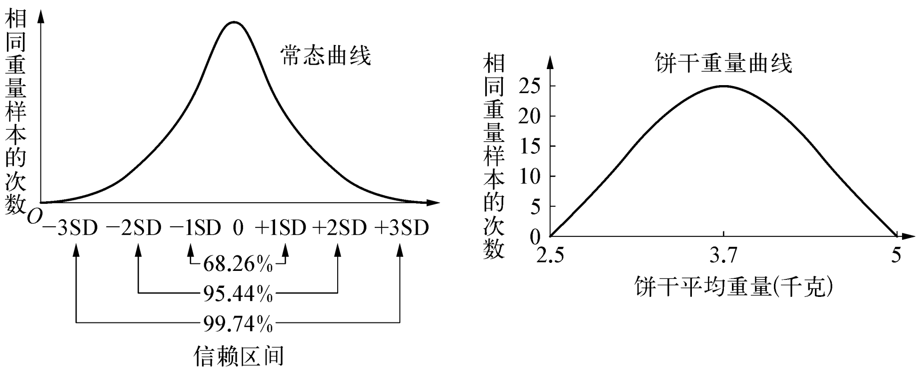


图 12-1 样本饼干信赖区间及重量曲线

(六) 样本平均数的抽样分配(the sampling distribution of the sample mean)

图 12-1 表示从 15 份饼干中抽选样本平均数的分布,此样本分布的标准差称为“标准误”(standard error)。标准误是一种用来显示样本结果反映真正总体参数值接近程度的统计测量,标准误愈小表示样本的平均数与总体的平均数愈接近。所以计算一个样本的平均数、变异数或标准差,这些用来叙述样本分布的值称为“统计量”(statistics)。如果样本是按概率法则选取的,则样本的统计量可被用来估计总体的“母数”(parameter)。

(七) 常态曲线(the normal curve)

当样本被重复使用或当样本量很大时,样本平均数的抽样分布趋近于“常态曲线”。研究者可利用常态曲线,并以概率理论为基础计算样本统计量与总体平均数的距离。标准误受母数变异数及样本大小所影响,样本量愈大常态曲线愈接近平均数的抽样分配[此称为“中央极限定理”(central limit teory)]。因为每个样本平均数的变异数须除以样本大小,所以样本量愈大,变异数愈小;在一个抽样分配中,标准误随着样本增加而减小。

(八) 信赖区间(confidence interval)

信赖区间即指定一个范围,以便使样本平均数会有特定百分比的机会落入该区间内。例如智商测验中,平均值为 100,若标准差为 8 的话,则一个标准差单位的面积(68.26%的人)具有 100±1SD 的分数,亦即落入 92～108。两个标准差单位时(95.46%的人),其智商的分数在 84～116。若采用三个标准差单位时(99.74%的人),其智商分数一定在 76～124。95.46%、99.74%等皆为信赖水准,而此区间称为“信赖区间”。

第四节　抽　　样

抽样(sampling)主要可分为随机抽样与非随机抽样。

一、随机抽样(probability sampling)

决定抽样的好坏最重要的因素，就是抽样是否在随机的情况下进行，因为样本之所以能真正代表总体，其基本的理论是因为抽样是通过随机的原则进行，所以有关随机的理论与抽样分配的原则十分重要。因此，当研究者运用乱数表或电脑程序软件，按一定的概率抽取样本，使每一样本有均等的机会被抽中时，即可称为随机抽样。任何抽样，若其样本没有同样公平的机会被抽中，便不是随机抽样。随机抽样又可以分为以下四种。

(一) 简单随机抽样(simple random sampling)

这是随机抽样中最单纯的一个方式，但必须先获得总体的名单，并对总体相当有把握时才能进行。此种方法最常用的方式是先利用随机号码表(random numbers tab)(或称乱数表)先给总体的名单上的每个人一个号码，运用固定的号码表，先随机抽出某一页、某一号为起点，开始向右或向下抽取一个号码(若总体的个数为千个，则以 4 位为一组，若为万个，则以 5 位为一组，直到达到样本数为止)。此种方法的缺点是当总体的名单无法获得时，此法便不可行，因为研究者无法把号码附在每一个样本上。

(二) 系统随机抽样(systematic random sampling)

系统随机抽样又称间隔随机取样(interval random sampling)，亦指每隔某些数字(K)的人就取出一个样本来。如下列公式所示：

$$K=\frac{\text{总体个数}}{\text{取样数目}}$$

除第一个采用随机方式抽出之外，以后每隔固定间隔(K)再取样，因此并不是每个人都有被抽中的机会，但大致而言，其机会还算公平。利用此方法时需注意避免一个现象：取样时，单位的次序不能和抽样间隔一致(例如，总

体的名单是某中学全校各年级、各班三名学生的总和,但是这三名学生则是按成绩优等、中等及劣等次序排列,如此一来,计算间隔数字时,一定不能是“3”,否则每隔三名取一名时,所取到的样本都是清一色全班最优的或是中等或最劣的,无法有一个包含各班、各种成绩的学生的客观样本),如此抽样就会产生系统偏差。

（三）分层随机抽样(stratified random sampling)

简单随机抽样较客观,但是因总体名单的困扰,在研究的实务中有时较难被使用,较常用的反而是分层随机抽样或整群抽样。分层随机抽样的步骤如下。

1. 首先研究者需将不同种类的人界定出来,每一类都必须明确,而且类与类之间必须有显著的互斥性,以便样本都只能归在其中一类。说明：研究者拟进行“不同职业的家长对给子女零花钱的态度差异研究”,拟进行分层随机抽样;“分层”就是把样本先分类,如把受访者的职业归类为公教类、农业、劳工类、商业类以及其他类等。这五类的定义必须清楚,彼此不能重叠。

2. 经过归类以后,每一类别内的样本彼此之间相似性很大,也因此研究者在每一类别中选取较少数的样本就足具代表性,取样的效率就可以增加。说明：类别与类别之间当然差异性很大,但是同类的样本其相似性越高,在抽样时,抽样效率就可提升。

3. 抽取的人数,最好能按人数比例来算。若总体的人口资料中,这五类人口占总人口的比例为1∶2∶1∶1.5∶2.5,则分层抽样的数字最好也符合这个比例,这样整个样本的结构与总体的结构差异不大。但事实上,我们关心的是每一层的不同情况,而不是总人口的问题,因此取样有时可以不照比例。但若使用不同比例的抽样时,在结论的解释时必须权衡人数的特性。

4. 运用分层抽样法时必须兼顾以下因素。

（1）分层的资料必须能获得,若各类层资料缺乏,何来分层随机抽样?

（2）每一分层中要具有相似性,否则被抽取的资料便不能代表该层的样本。

（3）需减少取样范围,以降低收集资料的费用,若费用太多,便失去了采用分层随机抽样的意义。

（4）要调查的情形最好不要超过一个以上的变量,以免造成分类的困难。例如,此分层抽样是研究不同行业的家长对给子女零花钱的差异性(理论上不

同行业的家长对如何给子女零花钱在态度上应有其差别)，但若研究者还要研究如何与亲子沟通时，以父母亲的行业类别来看，其差异可能不及以父母亲的教育程度来分析，但分层抽样时早就以行业类别来分，如此一来，抽样的方法与变量之间的关系就已失去了分层抽样的好处。所以，使用分层抽样时，要调查的变量最好不要超过一种。

（四）集丛取样(cluster random sampling)

当调查一个人数众多且各个区域分子都很复杂，异质性较高的总体时，集丛抽样是很好的抽样方法。

1. 先将总体分类或分层

例如研究者要在某市进行婚姻满意度的调查，某市有八个行政单位，其中五个在市区，三个在郊区。但基本上，这八区并非截然不同的景观，而且各区也都显现出复杂的人文景观，我们可以把每一个区都当作一个集丛。从抽样的角度来看，并不需要每个区都调查，应该从这八个区中随机取三个区或四个区即可。为周全计，研究者拟从五个市区中，集丛抽样选两个区，再从三个郊区中集丛抽样选一个区。万一总数还是太大的话，每一个集丛里又可以再做集丛抽样。例如，每一个区内再以街道为单位进行集丛抽样(前提是每个街道都是综合的街道，街道内的异质性也颇高时才可用集丛抽样)。

2. 扩大每一集丛的选择数

把每一个集丛(区)再划分为几个街道，再做集丛抽样，如抽出市区的几个街道，这些街道应可代替全部的街道了。

3. 分成集丛时，最好每个集丛之“内”愈复杂愈好

与分层抽样不同，分层抽样时，每个“层”里面最好同质性愈高愈好；而集丛抽样时，最好每个集丛的异质性愈大愈好。因为每个集丛的异质性愈大时，其代表总体的可能性就愈高。

二、非随机抽样(non-probability sampling)

非随机抽样中，每个人被抽到的机会是不完全相等的，每个样本不是有绝对一样的概率被选为研究样本。这种抽样方法适用于独特性的研究资料中，其样本无法在随机取样中获得。非随机抽样有以下五种类别。

（一）具备样本抽样(availability sampling)

当研究所需的样本早就在研究者的手中时，研究者不必再劳师动众地进行随机抽样的手续，所研究的样本可称为具备样本抽样。具备样本抽样又可称作偶然抽样或临时抽样(accidental sampling)，因为其样本是研究者随时可得的。具备样本抽样的指标很容易说明与描述，系针对某特殊行为的案主群体所做的研究，例如在医院服务的医务社会工作者以所服务的科别为样本做研究，研究者不必大费周折地从全国各地的医疗单位中从头做简单随机抽样，这种样本可称为具备样本抽样。

（二）配额抽样(quota sampling)

配额抽样是根据某些标准将总体分组，然后用非随机的方法来抽取样本，直到额满为止。例如，研究者进行学期制度修正与否的研究，因此对某大学的学生进行调查，以了解他们是否想改变一个学年两学期的学期制或一个学年三个学期的学期制。研究者不管当时学校各年级学生的数额，直接就决定各年级不分男女，不分系别，不管各年级、各科系人数如何，就决意以周三早上来校上课的学生为主，每年级各访问一百人作为研究的样本，此即为配额抽样。配额抽样的样本是指依据某些既定的标准或特色来抽样(如以周三早上来上课的各年级学生)，其样本之特色、数目均由研究者主观来决定，当然容易产生误差。

（三）立意抽样(purposive sampling)

研究者依据自己的研究目的及对总体的了解来选取样本。例如，“披头士”(Beatles)歌迷的特征研究，研究者主观确定谁是披头士歌迷，此称为判断性样本(judged sampling)。如果我们有足够的知识，要选出具有代表性的人选时，立意抽样经常被使用。例如，若以某市青少年的犯罪状况与国际上青少年犯罪状况做比较时，样本的选择便煞费周折，因为研究必须考虑代表性的人选，更须考虑样本在不同时间、不同空间所产生的差异，更需考虑不同领域的个案工作者对犯罪青少年的定义差异。如此一来，要克服这些研究的障碍就必须借立意抽样来选取样本，以便在解释上较为客观、合理。而对于易产生的“选择性判断”错误，只好借着多选择具代表性的人选，以他们彼此间的差异来抵消了。

（四）滚雪球抽样(snowball sampling)

此方法多运用在对某一特殊人口只熟知某一少部分人的情况下。该方法

可从已知的人数中去收集资料，并请他们介绍其周遭朋友或其他可能适合接受访问的案主。因此，样本是从少数中累积循环，一直到受试者全部调查完毕，或样本数已达到研究的要求为止。这一方法可用在较不严密或只具粗略计划的研究中，可采用其作为补充性的说明资料。

（五）便利抽样（convenience sampling）

便利抽样的资料获取是即时的，看到什么就用什么，手边有什么就使用什么。这种抽样的方式固然方便，但是其客观性值得怀疑。在社会工作研究的过程中，非不得已还是尽量少采用此方法为宜。

三、样本大小的问题（sampling size）

若是抽样为非随机抽样，研究者可以不考虑样本大小；但若为随机抽样时，则需考虑。样本的大小，主要视总体的特性及研究问题的本质而定，若总体的变数愈大、异质性愈高，则抽样的数额必须较大一点。反之，若样本的同质性很高，在抽样的数字上可以将就一点。通常以下公式可计算样本的大小：

$$n=\frac{S^2}{(S.E.)^2}$$

（S 为标准差，$S.E.$为标准误）

要计算样本到底要多少的步骤如下：

1. 须先知道总体的标准差是多少（假设得知总体的标准差是 1.4）；

2. 要决定我们能够忍受的标准误可以有多大（假设我们可容许的标准误为 0.012），则样本数的计算为：

$$n=\frac{(1.4)^2}{(0.012)^2}=\frac{1.96}{0.000\ 144}\approx 13\ 611$$

经此计算得知，研究所需的样本数约为 13 611 个。

这个算法乍看之下清楚明了，反正按着公式计算得到的样本数，谁也没话讲。不过这个算法最大的问题是：谁来决定标准误的大小，谁又能知道多少标准误是可以容许的范围？可能还得由有经验的研究者根据研究的特质、样本的特性来“拟订”标准误的大小，但拟订来拟订去，说不定到时却又只根据“经验”来做决定，如此一来，抽样又失去了客观性。

除了用标准误容许度来决定样本的大小外,还有其他一些方法可供参考。

1. 原则上总体的同质性高时,样本数可略低;反之,同质性低时,样本数应调高,这是抽样的最基本通则。

2. 看研究的主要变量有几个来决定样本数,一般以 1∶20 的比例来拟订样本数。但研究变量有几个有时也很暧昧,是重要的自变量与因变量才算数,还是连属性变量(如性别、年龄、教育程度、住宅区域等变量)也算数?

3. 若用回归分析来分析资料时,则用回归分析中到底有几个变量来决定样本数,一般以 1∶40 的比例来拟订样本数。若有五个自变量,加上一个因变量的话,那应有的样本数大约就是 240。之所以如此拟订,想必是考量到统计分析时必须有足够的样本,否则若样本数过低,就算得到很大的变异数(variances),你也不能说是样本之间的差异所致。

总之,研究是否准确,研究结果是否可以推论,取决于抽样是否正确。而抽样是否正确,首先就要看样本数是否合理。样本数固然可以引用公式来计算,但其确定,更取决于研究者的研究态度是否严谨。

四、其他问题的考量

(一)推论时不要言过其实

当研究经过抽样,且就样本做好分析时,就算有了显著的研究结论,但是这种结论只能推论到合乎抽样架构条件的对象上,因为研究的样本是根据某种架构而来,推论时,也应以同样架构的现象为宜。实际做研究时,有时会碰到一些在抽样架构中被忽略的对象,如名字未列于名单上,或家中未安置电话等。此时,研究者应视实际状况,仔细推断其影响,或评估其特质或反应,以使结论客观、合情、合理。

(二)提防抽样误差与非抽样误差

研究一用到抽样,除了要注意抽样所带来的误差外,非抽样误差(non sampling errors)亦不可疏忽。非抽样误差是指在收集或处理资料过程中产生的误差。如反应错误、偏见,受访者对问题不清楚、拒绝参与,或抄写错误等,这些误差并非因抽样的过程所导致,故称为非抽样误差。

(三)随机分派(random assignment)

random sampling 和 random assignment 常被混淆。random sampling 是

随机抽样，指一种抽样的过程。而随机分派（random assign）则是指将受访的个人随机地安排到一些实验的情况或团体中。在随机分派的过程中，每个人都有个号码，并以随机的方式被安排在适当位置或组别，以避免调查的偏见。在随机分派的过程中，对特殊样本的了解相当重要，若所调查的样本数本来就不多，且这些样本又具有极为特殊的变量影响力时，研究者就必须研究这些特殊变量所带来的“不正常信息”，并予以修正，否则就会失去随机的意义，进而影响到研究的结果。

第五节　结　论

抽样是研究法中相当重要的一环，若研究的抽样工作有了瑕疵，则整个研究的价值便大打折扣。在社会科学的研究中，完美的抽样相当难得，但是整个抽样的步骤、分析以及不足，都应在研究中交代，以便使读者在研读报告时知道如何判断研究的信度与效度，也能对研究报告的发现有其立场。此外，抽样的工作甚至牵涉到研究伦理的问题，研究者不可不慎。

关键名词

元素（element）：研究资料收集与分析的基本单位。

总体（population）：系固定研究单位与元素的集合体。

抽样架构（sampling frame）：根据研究主题与对象的范围，列出一份包含所有合乎抽样条件者的名单，再从这全部的名单中决定抽取百分之多少的计划，此为抽样架构。

抽样单位（sampling unit）：不同抽样阶段中的抽样元素。

抽样误差（sampling error）：总体的实际值与样本统计数估计值之间的误差值为抽样误差。

非随机抽样法（non-probability sampling）：非随机抽样法中，每个人被抽到的机会是不完全相等的，每个样本不是有绝对相同的概率被选为研究样本。

信赖程度与信赖区间（confidence levels and confidence intervals）：估计总

体的实际值时所制订的误差判定标准为信赖程度(alpha值)。而信赖区间则是总体的实际值在某一概率估计范围内数值分布的区域。

研究母体(study population):样本实际被选取架构中所有元素的集合体是研究母体。

随机抽样(random sampling):按一定的概率抽取样本,使每一样本有均等的机会被抽中时,即可称为随机抽样。

习　题

1. 试述立意抽样的优点及缺点。
2. 在选取随机样本时,一般可分为随机抽样及非随机抽样两类,又各有四种不同的抽样方法。请从这两类中各列举一种抽样方法,说明其意义、使用时机、优点及其缺点。
3. 何谓简单随机抽样? 如果你要从一个1 000人的总体中选出100人的简单随机样本,你会使用什么方法?
4. 试述分层随机抽样的适用时机,以及应用时的注意事项。
5. 如果由于经费的限制,自己无法进行抽样调查来获得研究所需的资料,你要怎么办?
6. 假设你拟从事某地区老人福利需求的调查研究,试回答下列问题:
   (1) 此项调查研究能让你达到什么样的目标?
   (2) 本研究的总体如何确定?
   (3) 你将采用的调查方法是什么? 为什么?
   (4) 如果采用抽样调查方法,你如何抽取样本(请提出抽样方法),此种抽样方法有什么优缺点?
   (5) 实地调查中,你将以哪一种调查方法(邮寄问卷法、访问法、电话访问法)收集资料,此种方法有何优缺点?
   (6) 从事此项研究,你必须考虑到哪些研究伦理的问题?

第十三章

资料分析方法

◎本章作者：邹平仪

第一节　前言

第二节　统计方法在社会研究资料分析中的功能、限制与选判原则

第三节　资料分析的统计方法

第四节　实务研究与资料分析的其他议题讨论

摘　要

资料分析为社会研究中的重要步骤之一。由于为探究社会现象所收集的原始资料往往蕴含着丰富的含义，因此，必须对问题本质和资料特征先行辨识，方能选择适当、正确的方法来分析、整理、验证资料，并且有效地解答问题，进而了解现象事实。

统计学是用经由系统性整理与数学演算后的数据资料来解读社会现象本质与样貌的学科。应用统计学方法虽然可以发挥其对于过去、目前、未来的实体情境进行描述、验证以及推估的功能，然而每一种方法也存在应用上的限制，倘若不当使用了某种统计方法，将可能误解数据的意义，甚至扭曲事实。据此，选择统计分析方法之前，宜先思考与判断下列几项因素：(1) 研究设计的变量属于何种测量层次；(2) 变量之间是对称的还是不对称的关系；(3) 统计量数是否具有消减误差比例的意义；(4) 研究的性质和目的是什么。

一般而言，社会工作研究中处理量化资料较常使用的统计方法包括下列五种。

1. 次数分配与百分比：用以简要描述研究对象(样本)各个变量特质的分布情形。

2. 标准分数之 Z 分数：用以描述个体(或某特定群体)某项行为表现在群体(或所有群体)中所在的相对位置。

3. 各类关联系数：用以检视现象事实之间所存在的关系是否具有统计学上的显著意义。其依据变量的测量层次、对称或不对称关系等特性，而选择适用的关联系数。在社会工作研究中，较常使用的关联系数包括：列联相关系数与卡方检定、皮尔逊的积差相关系数。

4. 各种平均数差异性检定：用以检视与考验现象事实之间所呈现的差异程度是否具有统计学上的显著意义。社会工作研究较常使用的差异性考验为平均数差异性检定，如 t 检定与 F 检定(变异数分析)。

5. 因果关系的回归系数：用以分析现象事实之间的单向线性关系。借由因果关系的探究，可以建立某一行为、现象的最佳预测模型。社会工作研究较常使用的因果关系分析为逐步回归与阶层式复回归分析。

在有关资料测量、分析过程中经常讨论的实务问题，如问卷设计的信度、效度检测、虚拟变量、误失值处理、样本数目估算标准等，亦于本章一并讨论。

第一节　前　　言

一般而言，实证科学的研究方式可以区分为两种类型：量化研究(quantitative study)与质性研究(qualitative study)。前者强调以标准而系统化的运作规则将研究概念、变量转化成可以测量、定义、操作的数值，经过计算、比较、验证的过程，将社会现象的特质予以通盘性描述，或经由变量之间关系的连动规则，来预测、推论社会现象的发展，此即通则取向(nomothetic approach)的研究形式。分析近五年硕博士学历者可支配所得的变化情形，探讨社工人员工作压力、因应策略与调适状况之间的关系，以及研究年长者健康状态的影响因素等议题均属此列。后者则是将特定主体或对象的叙说、观察、讨论收集而来的资料，借由过录、分类、命名、赋予概念、解说等抽丝剥茧的分析程序，而达到透析与诠释行为现象含义的方法。由于质性研究强调个别化及主观性，并且反对过度追求逻辑通则，因此被归属为独特取向(idiographic approach)的研究形式。几个实务应用上的案例如剖析受家暴妇女的心理复原历程、探究癌症晚期病友的自我价值定位，以及了解酒精成瘾者的戒断困境等议题即属此列。本章介绍的内容架构，主轴为量化研究的资料分析方法。

资料分析为社会研究历程中的重要步骤之一。其大多应用统计学原理，将原始资料，如问卷、量表、测验的测量结果，或访谈、行为观察记录等予以归类、简化，进而描述分析。又或者通过假设检验及总体值估计的运算，使人们对社会现象事实能有进一步的认识、了解与推测。由于社会研究所欲探讨的问题及所收集的资料往往包含着各种不同的性质与形式，因此，必须先对问题本质和资料特征加以辨识，方能选择适当的统计方法进而正确地分析资料，并且有效地剖析、解答问题。

统计学包含测度与量数的概念，通常是借由资料的收集、整理、陈列、运算、分析等过程，简要地呈现社会实情的样貌与社会现象的本质。通常来说，零散的原始资料(primary data)必须通过计算、测量、排列等方式重新汇整，才能使其成为有意义而易于理解的参考数据。所以，资料分析与统计分析(statistical analysis)两词常被混合使用。不过就名词的意义而言，前者较后者

广泛,但若从量化资料(quantitative data)为研究素材的主体而论,则后者较前者通用。

第二节　统计方法在社会研究资料分析中的功能、限制与选判原则

一、功能

统计方法经常被用以阐明自然及社会现象的一般性通则,也是管理、决策者分析状况与推测情境时所仰赖的重要工具。依据资料分析所能发挥的功能与其目的,统计学包含三大类别。

(一)描述

运用计算、测量、分组、排列次序等方法,将原始资料的属性特质进行陈述及说明,此即为描述或叙述统计学(descriptive statistics)。

(二)推论

由样本资料特质推估总体(群)的特征或分布状况,则为推论或推理统计学(inferential statistics)。一般包括以下几方面的主题。

1. 概率分配。

2. 抽样理论与抽样方法。

3. 推估与检定。

实务上因为对总体分配所能掌握的条件与程度不同,因此又将推论的范畴区分为母数及无母数统计学(parametric and nonparametric statistics)。

(三)实验

通过严谨的情境设计与样本分派程序,将受试者随机安排至不同的情境中(自变量)接受处遇,并控制其他条件使每位受试者的属性是一致的,而后对某一特定的行为现象(因变量)的变化加以观察、测量,此种研究方法称为实验设计(experimental design)。在实验过程中,通常分为有处遇介入的“实验组”与无处遇介入的“控制组”,实验组的受试者可能接受一种或一种以上的处遇内容,而后再比较实验组与控制组在标的行为(因变量)上的改变程度。因实验设

计在自变量的操作上可以有不同的形式，如重复量数、前后测或共变数分析(analysis of covariance)等，因此实验设计涉及较为繁复的概念，故而部分统计学者将其视为独立类别，笔者在此不再详细讨论。

二、限制

统计方法的应用虽可帮助我们解答生活上的问题，了解社会现象。然而若用之不当，则可能误解实况，只能得到失真的信息。因此，在使用之时宜留意以下原则。

1. 原始资料须先经过数值化的转换过程以及精确的测量程序，才能进一步分析处理。

2. 样本选取与样本大小的确定，都必须依照抽样理论的准则进行才能具有代表性。

3. 每一种统计方法均有其适用条件，故须先就资料性质判断之后再行选用。

三、选判原则

行为科学研究通常会将社会现象的概念转化为可测量的变量，而后收集与变量相关的资料，并使用统计方法测量及验证两两或多重变量之间的关系，以借之了解行为现象的通则、特性，进而推测其可能的发展趋势。以下即简要说明选择统计方法时需先行辨识的事项。

（一）变量的测量层次(level of measurement)

不同测量层次(类别、顺序、等距、比率或等比)的变量，须运用不同的统计测量方法，以适切分析变量特性及变量之间的关系(请参见第八章第三节“量度的性质”)。

（二）变量之间的关系是对称的(symmetrical)或是不对称的(asymmetrical)设计

变量之间如果呈现了自、因变量彼此影响、互为因果的形态，即属于对称关系；反之，倘若自、因变量显示了前因后果不可逆的次序形态，则属于不对称关系，即意味着因变量的表现受到了自变量的影响。这两种变量的关系形态须运用不同的统计测量方法。

(三)统计量(statistic)是否具有消减误差比例(proportionate reduction in error,简称 PRE)的意义

PRE 数值系表示以一个现象(自变量 X)来预测或解释另一个现象(因变量 Y)时所能够减少的误差比例(通常以百分比来表示)。由于探究行为现象关系的目的就是希望减少预测或解释时的错误,因此,具有 PRE 意义的统计测量法为研究者优先采用。

$$PRE=\frac{E_1-E_2}{E_1}$$

E_1:期望误差(expected errors),即假定不知道 X 的情况下预测 Y 的情况所产生的误差值。

E_2:观察误差(observed errors),指知道 X 的情况,而依据 X 的数值来预测 Y 的情况所产生的误差值。

由公式可知,X 与 Y 变量之间的关系愈强,所能减少的预测误差就愈大,因此,此消减误差值的大小,亦可以反映 X、Y 两个变量之间关系的强弱程度。

(四)研究目的与性质(the aim of social research)

根据社会科学研究的目的,大体上可将其分为探索性(exploratory)、描述性(descriptive)、解释性(explanatory)与解析性(analytical)等性质。依照各类型的研究重点而言,前两者以描述统计学的应用为主,后两者则属于母数估计与假设检定之推论统计学的范畴。

第三节　资料分析的统计方法

统计方法是资料分析过程中的必要工具,通常处理的是经由调查或观察收集而来的群体量数资料,其中非数值的部分亦须经由量度转换为可计算与测量的单位。目前社会科学中最常被用来运算大量数据资料的软件是 SPSS (Statistical Package for the Social Science)。相较于 SAS、BMDP 等其他统计软件,SPSS 的设计原理和操作指令较为简单易懂,尤其是中文版开发之后,其操作程序更加倾向以使用者为导向,因该软件并非本章撰述重点,因此请读者

依照实际需求自行参考相关书籍。以下即针对社会科学（社会工作）研究较常处理的资料性质、统计方法，摘要说明与示范。

一、样本特性分析（profile）：次数分配与百分比（frequency distribution and percentage）

将零乱分散的原始资料予以分类（组），并累计变量属性（类别）出现的次数，而后制作次数分配（frequency distribution）及百分比（percentage）图表，以达到简化资料、利于理解的目的。

（一）适用时机

描述研究对象（样本）属性变量（attribute variable）特质时使用。

（二）公式

1. f（次数）：变量内每一类属出现的数量。

2. p（比例）：$\frac{f}{n}$。

f：变量内各类属出现的数量；

n：样本总数。

3. PR（百分率）：

$$PR=\frac{f}{n}\times 100\%。$$

（三）范例

调查我国台湾地区 459 名各个领域社会工作人员的基本特质，如性别（sex）、年龄（age）、婚姻状态（mar）、宗教信仰（bel）、是否为本科毕业（swp）等现况的分布情形。

附注：459 名为邮寄 650 份问卷的回收总数，回收率为 70.6%

依照资料录入及电脑统计软件运算，可得到表 13－1 所示的结果。

表 13－1　受访者个人基本资料分析

变　量	类　别	人　数	百分比（%）
性　别	男	73	15.9
	女	386	84.1

(续 表)

变　　量	类　　别	人　数	百分比(%)
年　龄	20～29 岁	224	48.8
	30～39 岁	163	35.5
	40 岁及以上	72	15.7
婚姻状况	已婚	201	43.8
	未婚	258	56.2
宗教信仰	佛教	129	28.1
	基督教	77	16.8
	天主教或其他	56	12.2
	无	197	42.9
是否为本科毕业	是	344	74.9
	否	115	25.1

研究者依据表中陈列数据,简要描述受访社会工作人员的基本特质。

附注:有关 SPSS/PC+各程式指令语法、意义、范例及报表输出格式等详细讨论,另有专书详述,请自行参考:

王保进.1999.视窗版 SPSS 与行为科学研究.台北:心理出版社.

王丛桂,罗国英.1992.社会研究的资料处理.台北:黎明文化事业公司.

邱皓政.2009.量化研究与统计分析:SPSS 中文视窗版资料分析范例解析.重庆:重庆大学出版社.

Robert L. Miller, et al.. 2006. SPSS 在社会科学的应用.庄文忠,译.台北:五南图书出版股份有限公司.

陈正昌.2000.行为及社会科学统计学:统计软件应用.台北:巨流图书公司.

二、相对地位比较:*Z* 分数(*Z* score)

若比较不同群体在某一行为表现的差异,或者某特定群体在不同行为表现的优劣状况(相较于各行为表现的平均状态),标准分数(standard score)可以协助研究者处理此类问题。

（一）适用时机

1. 欲描述个体（或某群体）行为表现在群体（或全体）中的相对位置，以便了解个体（或某群体）的表现优于群体（或全体）表现（处于平均状态之上，相对表现较好）或劣于整体表现（处于平均状态之下，相对表现较差）。而这优于或劣于的差距，通过与平均数距离多少个标准差（standard deviation）单位来体现。

2. 比较不同群体（分组）在某一行为上的相对表现（优或劣）情形。

（二）公式

Z 分数是标准分数转换方法中的典型代表，其以平均数及标准差作为设算基准。因此，就行为现象测量资料的比较而言，Z 分数的应用确实达到了简明清晰的效果。此一统计方法，在描述与推论统计的研究范畴内均可运用。

$$Z=\frac{X-\overline{X}}{S_x}$$

式中，X 为某一观察值；$\overline{X}$ 为算术平均数（该群体）；S_x 为标准差（该群体）。

（三）范例

研究者调查各领域社会工作人员的离职率、工作满足感，希望能借以评估各领域社会福利机构的组织绩效及管理情形。现假设某县市政府、医疗、儿童、青少年、老人、身心障碍及妇女家庭等社会福利机构，其社工人员离职率与工作满足感的原始资料经由电脑换算处理后，得到如表 13－2、图 13－1 所示的结果：

表 13－2 各领域社工人员离职率与工作满足感的 Z 分数

工作领域	县市政府	医疗	儿童	青少年	老人	身心障碍	妇女家庭	平均数	标准差
离职率（%）	16.39	7.42	8.13	7.00	12.48	13.05	10.10	10.65	3.22
Z 分数	1.78	−1.00	−0.78	−1.13	0.57	0.75	−0.17	—	—
工作满足感	47.65	56.50	57.27	59.38	52.49	51.73	50.26	53.61	3.89
Z 分数	−1.53	0.74	0.94	1.48	−0.29	−0.48	−0.86	—	—

参照表 13－2、图 13－1 资料，可以得知下列信息。

1. 资料显示，除了妇女家庭类机构之外，其他各领域社工人员的离职率与

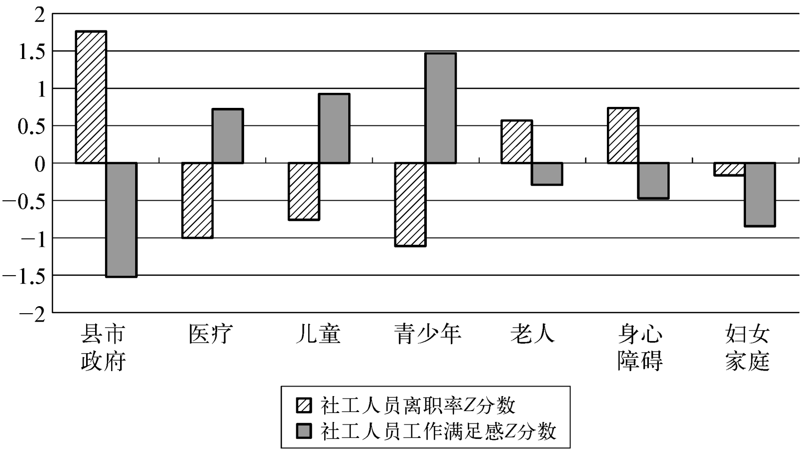


图 13－1　各领域社工人员离职率与工作满足感的 Z 分数分布

工作满足感均呈现反向关系,亦即除妇女家庭类机构外的社工人员的工作满足感越高,其离职率便越低,反之亦然。因此,假设离职率与工作满足感之间呈现显著关联的对称关系,机构管理者欲降低社工人员的离职率,则应从强化员工工作满足感的措施和策略上精进改善。

2. 服务于该县市政府与青少年工作领域的社工人员,其离职率与工作满足感呈现两种极端现象。前者离职率最高,工作满足感最低,后者则反之。站在机构管理者的立场,应对影响工作满足感和离职率的相关因素(如工作负荷量、组织气氛、薪资待遇、升迁机会、福利保障等情况)予以评估,而后再检视前述两个领域中的工作对应条件是否具有明显差异,如果事实确然,则前者应对后者的管理模式进行标杆学习。

附注:

1. 上述问题亦可利用单因子变异数分析(one-way ANOVA)的方法检测各服务领域社工人员离职率及工作满足感的差异情况是否达到假设检测上的显著意义,其后再决定是否需要进一步探究造成差异的原因。

2. Z 分数分布的条形图,可利用 Office、WPS 等软件绘制。

三、行为现象之间关系(关联性)的分析(correlation analysis)

社会科学研究大多以检视现象之间的关联性为主要目的,简言之,研究者可借由某一行为现象(X 变量)的表现,去预测另一个相关行为现象(Y 变量)的

状态或发展趋势。同理,研究者也可以借着改变一个可操作处理的行为,来影响另一个难以操作处理的关联行为。例如,以某人的教育程度(自变量)预测其薪资所得(因变量);以案主的高中成绩预测其高考得分;或者以人与人之间的互动频率预估两者的情感增进程度等。

各类相关系数是用来衡量两个变量之间关系强度的统计方法,兹将实务上较常使用的变量关联法列表如下(见表13-3)。

表13-3　双变数之测量层次及其适用的相关系数

Y	X		
	类别变量	次　序　变　量	等距或等比变量
类别变量	ϕ相关列联系数(C) Lambda(λ,对称关系) tau-y(不对称关系)	—	—
次序变量	theta(θ)	Spearman 等级相关(r_s) Kendall 和谐系数(计算不同评分者对相同样本一次或多次评定等级之间一致性的程度) Kendall tau-a, tau-b, tau-c Gamma(对称关系) d_{yx}(不对称关系)	—
等距或等比变量	点值二系列相关(r_{pb})	eta 相关比(η^2)(亦可借之检定变量之间是否存在曲线相关性)	Pearson 积差相关(r)

附注:高层次的变量可降低其测量水准(但会损失变量原始资料部分的数学演算性质)而采用较低层次的相关测量法。例如将年龄(等比变量)简化为老、中、青三组的次序或类别变量,以符合各种统计方法所要求的条件。

上述所陈列的关系测量方法,又以列联相关(contingency coefficient)与积差相关(product-moment correlation coefficient)系数最常为社会科学(社会工作)研究者所用,以下将分别说明。

(一)列联相关系数与卡方检定的显著性考验

列联相关系数是一种用来分析两个互相影响的类别变量关系强度的方法。由于系数本身的高低并不能代表两者相依共变的程度是否具有意义(显著关系),因此,列联相关系数多半与卡方考验一并讨论。

1. 适用时机

(1)列联相关系数

① X、Y 两个变量均属于类别测量尺度。

② X、Y 两个变量均包含两个以上的类别,如性别(男、女)与是否担任主管

职务(是、否),简而言之,至少是 2×2 方形列联表以上的变量关系分析。

(2) 卡方考验

① X、Y 两个变量均属于类别测量尺度,样本经由随机方式选取。

② 依资料特性与不同的分析目的,可以区分为四种常用类型。

a. 适合度考验(goodness of fit test)

适合度考验用来检定实际观察次数(或百分比)与某种理论次数(或百分比)的分配状况是否相符。此外,适合度考验同时也是检视样本观察的次数与理论或总体的次数分布是否具有显著性差异的分析方法。例如,某所大学的学生性别分布比例是否与某市人口性别比例(1∶1)相符。由于此种考验内容仅涉及一个变量,因此亦可视为单因子检定。

b. 独立性考验(test of independence)

独立性考验用于以下情况:自某总体抽取样本,欲了解经由调查而得的 X、Y 变量观察值其间是否具有关联性。例如,社会经济地位(X 变量分为高、中、低三类)与社会声望(Y 变量也分为高、中、低三类)之间是彼此独立的关系(不同的社会经济水平者,他们的社会声望评价并没有特别的差异现象),抑或是相互影响的对称关系(不同的社会经济水平者,他们的社会声望评价也不同,而且社会经济地位越高者,其社会声望的评价也越好;当社会声望的评价越好时,也会愈加提升其社会经济地位)。由于独立性考验乃是同时讨论样本群中两个变量之间的关联程度,因此属于双因子检定。

c. 百分比同质性考验(test for homogeneity)

自若干总体随机抽取样本,而后依据各样本群行为观察值的百分比分布,来考验此若干总体在某行为表现上的分布状况是否具有同质性。例如,考察三大宗教信仰团体(天主教、基督教、佛教)对医疗机构对癌症晚期病人执行“安乐死”的态度反应(赞成、不赞成)有无不同。同质性检定的目的,系为探讨不同样本群组(自变量)的某一行为表现(因变量)有无显著差异。在学理上,由研究者所操作处理的变量除了被视为自变量外,同时也应被视为设计变量,此外因变量又被称为反应变量。整体来说,百分比同质性考验是一种具有前因后果的不对称(asymmetrical)关系模型。

d. 改变的显著性考验:麦内玛考验(McNemar test)

检定同一组受试者经由设计,其行为表现(反应变量)在受试前后的改变程度是否具有显著意义。例如,60 位村民观赏核能建设成果影片前后,其赞成或

反对兴建核能发电厂的态度有无明显变化(改变)。

2. 公式

(1) 列联相关系数

$$c=\sqrt{\frac{\chi^2}{N+\chi^2}}$$

如为 2×2 以上的非方形(横行数与纵列数不同)列联表,可使用克瑞玛 V_c 统计数(Cramer's statistic)表示关联程度。

$$V_c=\sqrt{\frac{\phi^2}{\min(I-1,\ J-1)}}$$

附注:

式中,$\phi=\sqrt{\frac{\chi^2}{N}}$;公式的分母则指横行数$(I-1)$或纵列数$(J-1)$中自由度较小者。

(2) 卡方考验

① 适合度、独立性、百分比同质性考验

$$\chi^2=\sum\frac{(f_o-f_e)^2}{f_e}$$

自由度为

- 适合度检定 $df=k-1$
- 独立性检定 $df=(r-1)(c-1)$
- 同质性检定 $df=(r-1)(c-1)$

其中,　f_o为观察次数(observed frequency);

f_e为期望次数(expected frequency);

k 为单因子分数(组)的数目;

r 为因子分类,列的数目;

c 为因子分类,行的数目。

② 耶兹校正考验(Yate's correction for continuity)

在 2×2 的列联表(contigency table or crosstabulation)中,当期望次数介于5~10($5\leqslant f_e\leqslant 10$),即应使用本方法:

$$\chi^2=\sum\frac{(|f_o-f_e|-0.5)^2}{f_e}$$

③ 费雪正确概率考验(Fisher's exact probability test)

在 2×2 的列联表中,若期望次数小于 5($f_e<5$)或样本人数小于 20 时(SPSS 系统内定值),则应使用本方法进行次数资料(观察值)的独立性考验。

$$P=\frac{(a+b)!\ (c+d)!\ (a+c)!\ (b+d)!}{N!\ a!\ b!\ c!\ d!}$$

a	b	$a+b$
c	d	$c+d$
$a+c$	$b+d$	N

- a、b、c、d 为列联表中四个细格的观察值
- N 为样本数
- P 为该种特定排列组合的概率

④ 麦内玛考验

$$\chi^2=\sum\frac{(r-s)^2}{(r+s)},\mathrm{d}f=1$$

- r 指前测时态度反应为正向(如赞成),而后测时却改变为负向(如不赞成)的观察值
- s 指前测时态度反应为负向,而后测时却改变为正向的观察值

3. **范例**

(1) 列联相关系数与卡方考验(百分比同质性考验)

由于病人的生活自理能力与照顾者的压力负荷情形息息相关,因此,某医院社工想要了解病人疾病性质(癌症、非癌症组)与失能程度(重度、中度、轻度)之间的关系。换言之,社工想要探究癌症病人组(重症者)的自我照顾能力,是否与其他非癌症病人组(糖尿病、脑血管病变)的自我照顾能力有不同。经调查研究后,原始资料分析结果如表 13-4 所示。

表 13-4　病人疾病性质与失能程度的分析

因变量失能程度	自变量		卡方值(显著水准)	列联系数 V_c
	病人疾病性质			
	非癌症病人组 N(%)	癌症病人组 N(%)		
重　度	45(45.0)	19(21.1)	12.11**(0.002)	0.24
中　度	19(19.0)	24(26.7)		
轻　度	36(36.0)	47(52.5)		

$\mathrm{d}f=2$, ** $p<0.01$

卡方考验显示：病人疾病性质与其失能程度之间呈现显著关联性（$p<0.01$）。非癌症病人失能程度属重度者约占半数（45.0%），照顾者身心负荷可想而知。而癌症病人组失能程度则以轻度者居多（52.5%），相形之下，其存活期间为照顾者所带来的压力负荷，反而比非癌症病人的照顾者轻。

（2）麦内玛考验

某社会学家对某市某地区60位村民进行研究，欲了解施行认知训练（影片欣赏、参观、演讲）是否影响其对兴建核能发电厂的态度。经过调查，认知训练前后，村民的态度反应及麦内玛考验结果列表如下（见表13-5）。

表13-5　认知训练前后，村民的态度反应及麦内玛考验结果

		认知训练后		McNemar test 显著水准
		不赞成	赞成	
认知训练前	赞成	8	17	0.017 6*
	不赞成	13	22	

$df=1$，* $p<0.05$，双尾检定

上表资料显示，认知训练前后，村民对兴建核能发电厂的态度出现了显著性改变（$p<0.05$）。从列联表陈列的数据得知，村民态度由不赞成转变为赞成者（22人）相较从赞成转变为不赞成者（8人）为多。整体而言，认知训练使得较多的村民支持兴建核能发电厂。简言之，该项认知训练方案达到了改变态度倾向的预期目标。

（二）积差相关

皮尔逊（Pearson）积差相关系数经常被社会科学研究用来检验双变量之间的关联性。通常以r系数（分布范围介于±1）的大小代表双变量之间相互影响的程度，其意义包括了完全负相关（-1）、负相关、零相关、正相关及完全正相关（$+1$）等关系情况。

1. 适用时机

X、Y两变量均为等距或等比测量层次时使用。

2. 公式

$A.Z$值之积的总和平均数：

$$r_{xy}=\frac{\sum(X-\overline{X})(Y-\overline{Y})}{\sqrt{\sum(X-\overline{X})^2(Y-\overline{Y})^2}}$$

3. 范例

某研究者欲了解社会工作者个人的生涯规划程度与专业承诺感之间的关系。从 459 名受访社工人员的资料分析结果可知 $r=0.405\ 2$（$p<0.001$）。表示社会工作者的生涯规划程度越高，其专业承诺感也越高，两者呈正向的连动关系，且关联系数达到统计上的显著意义。

四、行为现象之间差异关系的假设考验（hypothesis testing）与平均数差异检定（difference-of-means test）

社会科学的研究常受限于某些现实条件的考量，譬如有限的人力、财力、物力、时间等因素，因而无法对所关心的现象事实进行全面性的普查。因此，必须借由严谨的抽样方法（请参考本书第十二章“抽样”内容说明），选取总体中部分个体（样本）进行研究，并依照假设考验原理，检定样本资料所呈现的特质是否能在某一个误差容忍水准之下推演论点，以协助我们了解整个较为普遍、广泛而可能存在的现象事实（总体）。此即为假设考验的意涵，前述卡方检定及积差相关所讨论的“显著性”等问题，也属本概念的延伸应用。一般而言，假设考验分为以下几个步骤。

（一）建立假设（making assumptions）：虚无假设（null hypothesis，H_0）与对立假设（alternative hypothesis，H_1）

例如，H_0：$\mu_1\leqslant\mu_2$（乡村居民生育率低于或等于都市居民生育率）；

H_1：$\mu_1>\mu_2$（乡村居民生育率高于都市居民生育率）（本例属于单尾考验）。

（二）选取适当的抽样分配（obtaining the sampling distribution）

例如：Z 分配、t 分配、χ^2 分配、F 分配等

（三）选择显著水准及临界区（selecting a significance level and critical region）

较常使用的显著水准为：

1. $\alpha=0.05$（以 * $p<0.05$ 表示）；

2. $\alpha=0.01$（以 ** $p<0.01$ 表示）；

3. $\alpha=0.001$（以 *** $p<0.001$ 表示）。

依照研究目的决定单侧（尾）（one-tailed）或双侧（尾）考验（two-tailed

test)。单侧考验是用来处理单一方向方面的问题，换言之当研究假设中含有“大于”“优于”“高于”或“小于”“劣于”“低于”的用语时，即须将所有的 α（显著水准）完全集中在常态分配图的右端（统计系数为正）或左端（统计系数为负）的一个区域进行决策判定。例如，研究假设 $H_1: \mu_1 > \mu_2$，即表示假定第一样本群组的平均数减去第二样本群组的平均数为正，此处是将 α 全部集中在常态分布曲线的右端、正值一边进行讨论（第一样本群组某行为的平均表现优于第二样本群组）。双侧考验则是一种不强调方向性、仅检定有无差异的研究设计。例如，$H_1: \mu_1 \neq \mu_2$，即表示研究者所关心的是两个样本群组行为观察值的平均数有无显著不同，其并不需要区分两者的高低或优劣，因此所拟订的 α 便须置于常态分布曲线的两端（$\alpha/2$）再执行决策判定，也就是同时检测常态分布曲线左右两侧临界区域的拒斥标准。

（四）计算检定统计量(computing the test statistic)

根据选用的统计量数方法（如 Z 值、t 值、X_2 值、F 值等）及样本观察数据资料计算其统计数值。

（五）做决策(making a decision)

依据原先所订的显著水准（α）来判断是否拒绝(rejector fail to reject)虚无假设（H_0），为假设检定做出结论。

论文与专题进行研究结果判定时，都需要依照假设检定步骤进行。因文章篇幅有限，所以研究者则多半将此陈述予以简化或省略。以下即就实务研究较常使用的差异性考验 t 检定与 F 检定（变异数分析）予以说明。

1. t 检定分析(t-test analysis)

t 检定是一种相当强韧的(robust)统计考验方法，一般被应用于两个小样本群平均数差异的检定，包含独立样本与相依样本（重复量数：repeated measures）两种特性的分析。t 分配的特质与 Z 分配相当类似，它的曲线分布亦呈平均数为 0，左右对称的形态，唯其峰度较高狭，曲线两端距底线较远。当样本数越大时（一般指 $N>120$），t 分配便越趋近于 Z 分配（请参阅各统计学书目附录部分，比较 Z 分配与 t 分配的临界值标准）。由于社会科学研究很难完全掌握总体的性质（总体标准差 σ 已知的条件下，可使用 Z 检定法），因此，此处仅就 t 检定分析的概念予以说明。

1）适用时机

(1) 基本假设为两个小样本群平均数属常态分配，变量测量层次为等距、比率尺度。

(2) 小样本之平均数差异检定：单一样本群时，指 $N \leqslant 30, N = 50$；两个样本群时，指 $N_1 + N_2 < 120$。

(3) 当总体的标准差(σ)或变异数(σ^2)未知时使用。

2）公式

(1) 单一样本时(考验样本平均数与总体平均数的差异是否显著)

$$t = \frac{\bar{x} - \mu}{\frac{\hat{S}}{\sqrt{n}}}，其中 \hat{S} = \sqrt{\frac{\sum(x - \bar{x})^2}{n - 1}}（总体标准差的不偏估计值）$$

$$df = n - 1$$

(2) 两个独立样本时

需先考验两个总体变异数是否相等($\sigma_1^2 \neq \sigma_2^2$ 或 $\sigma_1^2 = \sigma_2^2$)，即检定 F 值是否达到显著水准。F 值系以大的样本变异数($\hat{S}_L^2$)除以小的样本变异数($\hat{S}_S^2$)。

① 当 F 检定达显著水准时($p < 0.05$)，代表两个样本群所来自的总体变异数不相等($\sigma_1^2 \neq \sigma_2^2$)，宜采用 Cochran 及 Cox(1950)所创的分隔变异数 t 检定(separate variance t test)。

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\hat{S}_1^2}{n_1} + \frac{\hat{S}_2^2}{n_2}}}，df = \frac{\left(\frac{\hat{S}_1^2}{n_1} + \frac{\hat{S}_2^2}{n_2}\right)^2}{\frac{\left(\frac{\hat{S}_1^2}{n_1}\right)^2}{(n_1 - 1)} + \frac{\left(\frac{\hat{S}_2^2}{n_2}\right)^2}{(n_2 - 1)}} \cdots 四舍五入取整数$$

② 当 F 检定未达显著水准时($p > 0.05$)，代表两个样本群所来自的总体变异相等($\sigma_1^2 = \sigma_2^2$)，此时则应采用合并变异数 t 检定(pooled variance t test)。

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\hat{S}_P^2}{n_1} + \frac{\hat{S}_P^2}{n_2}}}，其中 \hat{S}_P = \frac{(n_1 - 1)\hat{S}_1^2 + (n_2 - 1)\hat{S}_2^2}{n_1 + n_2 - 2}$$

此为合并变异数，即个别变异数的加权平均，$df = n_1 + n_2 - 2$。

研究者应依据变异数同质性考验结果，选择分隔变异数或是合并变异数统计量来研判检定结果。

③ 相依样本重复量数时，即相同受试者于不同时间（通常是处遇前与处遇后）接受了两次观察测量，研究者希望了解服务方案是否有效改善受试者的特定行为，此时应采用下列 t 统计检定。

$$t=\frac{\overline{D}-0}{S_{\overline{D}}}=\frac{\overline{D}+0}{\frac{S_{D}}{\sqrt{N}}}=\frac{(\bar{x}_1-\bar{x}_2)-0}{\sqrt{\frac{\sum D^2-\frac{(\sum D)^2}{N}}{N(N-1)}}},\mathrm{d}f=N-1$$

式中，N 为人数；$D=x_1-x_2$；$\overline{D}=\frac{\sum D}{N}=\bar{x}_1-\bar{x}_2$；分母为 $\overline{D}$ 的标准差，即 $\hat{S}_{\overline{D}}$。

3）范例

（1）两个独立样本的 t 检定

某医院社会工作部门欲评估出院准备服务——“早期干预”辅导方案对于缩短病人住院天数是否具有效果。研究者以系统性随机取样方法，依照科别比例抽样纳入 347 名病人，并将其随机分派至实验组（接受社会工作者辅导）和控制组（不接受社会工作者辅导）。辅导方案结束后，以 t 检定法比较两组病人平均住院天数，检定结果如表 13-6 所示。

表 13-6　出院准备服务与病人住院天数的 t 检定分析

组　别	人数	平均数	标准差	自由度	t 值	p 值
实验组	181	−5.88	15.21	329.9	−2.17	0.03*
控制组	166	−2.79	11.20			

* $p<0.05$

附注：平均数为负数者，指病人住院天数相较于该疾病标准住院天数之差值的平均数为负数，亦即病人的平均住院天数比公认标准短。

两组病人平均住院天数经 t 值检定达到显著差异（$p<0.05$，单尾检定，$\sigma_1^2\neq\sigma_2^2$），实验组病人的平均住院天数较控制组短了 3.09 天，表示该项辅导方案具有

缩短病人住院天数的效果。

(2) 相依样本(重复量数)的 t 检定

某青少年辅导机构欲评估其辅导方案对于改善虞犯青少年自我概念的效果,该研究使用人格测验量表(已建立信、效度)作为评量自我概念的工具,并于辅导完成前后分别对 10 名虞犯青少年进行观察测量。两次评量分数经由 t 检定得到下列结果(见表 13－7)。

表 13－7 青少年辅导方案对改善其自我概念的 t 检定分析

组 别	人数	平均数	标准差	自由度	t 值	p 值
辅导前—辅导后	10	4.70	0.92	9	5.11	0.001***

*** $p<0.001$

表中资料显示,虞犯青少年于接受辅导前后的自我概念评分具有显著性差异($p<0.001$,双尾检定)。即受试者经由专业辅导之后,其自我概念平均分数显著提升了 4.70 分。这也表示该项辅导方案具有持续执行与推广的价值。

2. 变异数分析(analysis of variance,简称 ANOVA)

变异数分析由统计学家费雪(R. A. Fisher)创立,其主要功能是检定多个样本群所代表的总体平均数之间(μ_j;$j=1,2,3,\cdots,k$,亦即 k 组总体的各个平均数)是否具有显著差异。统计学上一般以变异(离差平方和,SS:sum of squares)表示差异状况,而变异通常源自两方面的原因:一为已知原因而形成的变异(如处遇方案);二为抽样误差及未知原因而形成的变异(或称为实验误差)。当研究者欲分析两组以上平均数的差异时,利用 t 检定之两两组合的方式较为繁复不便,而使用变异数分析法(F 检定)则可达到化繁为简的效果。亦即先行检定 F 值是否达到显著水准(总变异量是否具有显著意义),如果总变异量具有显著意义,则进一步再行确认群组差异的来源。亦即应用事后比较(a posteriori comparisons)方法检视差异发生于哪些样本群之间。变异数分析的类型颇多,此处以较常使用的单因子变异数分析(one-way analysis of variance)为说明范例。

1) 适用时机

(1) 自变量为类别或顺序尺度,因变量为等距或比率尺度之平均数差异的

显著性检定。通常用于检定自变量类别在三组或三组以上的总体平均数差异性考验。

（2）样本来自总体，须符合 ANOVA 基本假设，即各个样本群的总体为以下几种情况。

① 常态性（normality）。即呈现常态分布。

② 独立性（independence）及可加性（additivity）。指变异来源甚为明确，分割后的各组群是彼此独立或不相重叠的。

③ 变异数相同（homogeneity of variance）。当样本资料经过考验不符合此项假设时，则须应用转换方法将其处理为变异数较为接近的情况，才能进一步使用本方法分析资料。

2）公式

（1）各组人数相同时

① SS_t（总变异量）

$$=\sum\sum(X_{ij}-\overline{X})^2$$

$$=\sum_j\frac{(\sum_j X_{ij})^2}{n_j}-\frac{(\sum\sum X_{ij})^2}{N}$$

式中，N 为各组样本数之加总，即代表总样本数。

② SS_b（组间变异量）

$$=\sum\sum(X_j-\overline{X})^2$$

$$=\sum_j\frac{(\sum_j X_{ij})^2}{n_j}-\frac{(\sum\sum X_{ij})^2}{N}$$

$$=\left[\frac{(\sum X_{i1})^2}{n_1}+\frac{(\sum X_{i2})^2}{n_2}+\frac{(\sum X_{i3})^2}{n_3}+\cdots+\frac{(\sum X_{ik})^2}{n_k}\right]-\frac{(\sum\sum X_{ij})^2}{n}$$

此处 $n_1=n_2=n_3=\cdots=n_k$（各组人数相等）。

③ SS_W（组内变异量）$=SS_t-SS_b$ 或$=\sum\sum X_{ij}^2-\sum_j\frac{(\sum_j X_{ij})^2}{n_j}$

④ MS_b（mean square，均方为组间变异数的不偏估计值）

$$=\frac{SS_b}{K-1}$$

式中，K 为自变量所区分之组别数，$K-1$ 为组间变异的自由度。

⑤ MS_w(组内变异数不偏估计值)

$$=\frac{SS_w}{N-K}$$

式中,$N-K$ 为组内变异的自由度。

⑥ F(F 值)

$$=\frac{MS_b}{MS_w}$$

查表之临界值,横行为分子的自由度($K-1$),纵列为分母的自由度($N-K$)。

(2) 各组人数不等时

计算方式与②完全一致,但须注意 SS_b 中 $n_1 \neq n_2 \neq \cdots \neq n_k$。

(3) 重复样本时

相同的样本接受不同的实验处理(即社会工作指称的处遇方法),用以检定或评量各种实验干预是否具有显著效果。此种设计必须将受试者个别差异所造成的误差从组内变异数中扣除,再计算所剩余的残差误差(不能确认来源或不能解释的误差),而 F 值便是由实验处理的变异数不偏估计值与残差变异数不偏估计值的除数所构成的。

$$F=\frac{SS_{\text{b. treatment}}/(k-1)}{SS_{\text{residual}}/(n-1)(k-1)}=\frac{MS_{\text{b. treatment}}}{MS_{\text{residual}}}$$

由于实务上极少为了评估比较(重复量数)各种处遇方法的辅导成效而让同一服务对象接受多种不同处遇方法,因此此检定方法较少被研究者采用。

(4) 事后比较

当 F 值到达显著水准,须进一步检定差异发生于哪些样本群之间(检定各组平均数之间的差异)时,即用事后比较法。以下即分别说明较常使用的事后比较法(通常为非正交比较法)。

① 薛费法(Scheffe method,S 法)

适用于各组人数不相等、两两比较包含两个以上组别之平均数差异的情形。各组人数相等时亦可使用。

$$F=\frac{(\overline{X}_i-\overline{X}_j)^2}{MS_W\left(\frac{1}{n_i}+\frac{1}{n_j}\right)},\mathrm{d}f=(K-1),(N-K)$$

其中$\overline{X}_i$与$\overline{X}_j$为两两相互比较之各组别的平均数；n_i与n_j为两两比较的各组样本数；K为组数；MS_W为误差均方和，也就是组内变异数不偏估计值。

计算所得的F值须大于查表临界值，方可裁决组间差异达到统计检定上的显著水准。

② 杜凯氏法（Tukey's honestly significant difference，HSD 法）

适用于各组样本人数相等、两两组别之间平均数差异的比较。

$$q=\frac{\overline{X}_L-\overline{X}_S}{\sqrt{\frac{MS_W}{n}}},\mathrm{d}f=K,N-K$$

其中，$\overline{X}_L$为所有组别中最大的平均数；$\overline{X}_S$为所有组别中最小的平均数。

或当各组人数不相等时，公式改为：

$$q=\frac{\overline{X}_L-\overline{X}_S}{\sqrt{\frac{MS_W}{2}\left(\frac{1}{n_L}+\frac{1}{n_S}\right)}},\mathrm{d}f=K,N-K$$

当q值大于$[q_{1-\alpha,(K,N-K)}]$查表临界值，则表示两组平均数的差异达到显著水准。

③ 纽曼-柯尔氏法（Newman - Keuls method，N - K 法）

适用于各组样本人数相等的差距考验，其最大特色乃是依照各组平均数之大小次序使用不同的临界q值，因而敏感度较高，比 HSD 法容易达到差异考验的显著水准。

$$q=\frac{\overline{X}_L-\overline{X}_S}{\sqrt{\frac{MS_W}{n}}},\mathrm{d}f=r,N-K$$

r为两两组合比较的两个平均数在平均数排列次序中相差的等级数（the number of steps）；计算而得的任何一组平均数差异值（$\overline{X}_L-\overline{X}_S$）必须大于该$q_{1-\alpha(r,\ N-K)}\sqrt{\frac{MS_W}{n}}$，才表示达到了显著水准。

3）范例

某医院社会工作者欲了解慢性病人自我照顾能力是否对其照顾者的身心

负荷造成了不等程度的压力。于是研究者对院内190名相关病人的自我照顾能力进行评估与调查,并借由压力量表测量其主要照顾者的压力知觉程度。研究资料经过单因子变异数分析后得到表13-8所示的结果。

表13-8 病人自我照顾能力及其照顾者压力之单因子变异数分析

变异来源	离均差平方和	自由度	均方	F 值	显著水准
组间(自我照顾能力)	307.1	2	153.6	3.66	0.028*
组内(照顾者压力)	7 836.5	187	41.9		
全体	8 143.6	189			

* $p<0.05$

表13-9 薛费法事后比较

自我照顾能力	平均数	完全无法自理 39.9	部分自理 39.4	完全自理 37.2
完全无法自理	39.9	—	0.355	0.031*
部分自理	39.4		—	0.058
完全自理	37.2			—

* $p<0.05$

根据单因子变异数分析结果可知,病人的自我照顾能力对其照顾者压力负荷程度造成显著影响($p<0.05$)。而由薛费事后比较则可进一步了解,此一差异主要来自"完全无法自理"与"完全自理"两组病人之间($p<0.05$)。前组的照顾压力(39.9±6.5)显著高于后组(37.2±6.1)。此研究结果再次印证了邱启润(1987)及钟娑娑(1989)等的研究论点,即病人的自我照顾能力(对照顾者的依赖程度)是照顾者身心压力的重要影响因素(陈光耀等,1992)。

五、行为现象之间的影响性分析:逐步回归(stepwise regression analysis)与层级式复回归分析(hierarchical multiple regression analysis)

回归与相关分析之间的差别是前者用以分析自、因变量之间的不对称关系(asymmetrical relationship),而后者则是用以分析自、因变量之间相互影响的对称关系(symmetrical relationship)。当行为现象通过研究观察而得知其间的连动关系与影响情况时,便可以借由了解属性变量、操作自变量而来影响或预

测另一个因变量可能的发展，此为回归分析法的应用条件与目的。回归分析通常以线性模型的回归方程式来呈现变量之间的关系。而复回归（multiple regression）的用途同上所述，主要是以一组自变量（X_1，X_2，X_3…）来预测、估计某个因变量（Y）。此方法可以计算出全部自变量对因变量变异量的解释程度（以 R^2 表示），以及每一自变量在预测因变量时的相对重要性（以 β_k 表示）。以下即针对社会工作实务研究中较常使用的逐步回归分析（stepwise regression analysis）与阶层式复回归分析（hierarchical multiple regression analysis）予以说明。

（一）逐步回归分析

逐步回归分析是将自变量逐步（一次一个，或一次数个）代入回归方程式中予以分析的一种复回归法，其对因变量变异量的解释力亦随之被逐一评估。资料处理时，自变量纳入分析的顺序系依照统计值决定，亦即由各自变量与因变量之间净相关系数的大小来决定各自变量被评估的顺序。净相关最高者优先选入分析，其余则重新计算净相关系数，再自其中选取系数最高者纳入分析。当其中部分相关的 F 值概率大于剔除概率（pout）时，则将该变量去除，而后持续依照上述步骤反复进行择入及剔除的过程，直至所有自变量已被纳入分析为止。据此，回归方程式所呈现的算式，应该是“以最精简的自变量来解释因变量的最大变异量”的最佳预测模型。

附注：逐步回归选取自变量的方法包括数种：顺向选择法（forward，亦称逐步引入法）、反向淘汰法（backward，亦称逐步剔除法）及逐步回归法（stepwise，亦称有进有出法）。而前面所述的选项、除项反复过程的做法，系指较常被引用的逐步回归法。

1. 适用时机

（1）通常使用于探索性研究。目的是从多个自变量中找出重要而具有意义的预测变量，并淘汰其他没有影响力或预测力的变量，而新的研究假设亦能借此建立。

（2）自、因变量的测量层次均属等距或比率变量（自变量如是类别或顺序变量，均须转换成虚拟变量后方能加以分析）。

2. 公式

多元回归分析的计算过程十分复杂，观察资料通常需要借助电脑软件处

理。以下仅就临床应用上较重要的概念、公式进行说明。

(1) 多元回归方程式撰写模型

利用“原始分数”计算的方程式为:

$$\hat{Y}_i = b_1 x_1 + b_2 x_2 + \cdots + b_p x_p + a$$

利用“Z 分数”或“相关系数”计算的方程式为:

$$\hat{Z}_Y = \beta_1 Z_{x1} + \beta_2 Z_{x2} + \cdots + \beta_p Z_{xp}$$

由原始分数所撰写的回归方程式可以利用观察值来计算并预测因变量 f 的情形。基于自变量所采用的衡量单位可能不同,如以日、月、年计算出的 b(斜率)值并不相同,因此,不能以 b 的大小直接判断其对因变量变异量的影响与解释程度。然而,倘若使用经过 Z 分数处理后的标准化回归方程式,则可依据转换后的标准化斜率 β(beta)系数的大小来直接判断各个自变量对因变量的相对重要性(β_1,β_2,…,β_p表示自变量对因变量变异量解释的加权数)。

(2) 回归系数的考验

用以检定效标变量(因变量)Y 与每一个预测变量(自变量)X 之间是否存在显著的直线关系,亦即考验每一个自变量的个别斜率(回归系数)是否具有统计学上的显著意义。

$$t = \frac{\hat{b} - 0}{\sqrt{Var(\hat{b})}}$$

其中,$\hat{b} = \dfrac{\sum (X - \overline{X})(Y - \overline{Y})}{\sum (X - \overline{X})^2}$

$$Var(\hat{b}) = \frac{\sigma^2}{\sum (X - \overline{X})^2}$$

$$\sigma^2 = \frac{\sum (Y - \hat{Y})^2}{n - 2}$$

(3) 净相关(partialcorrelation)与部分相关(partcorrelation)

净相关是利用数学方法把与两个变量有关的其他变量的影响力排除之后,

再计算这两个变量之间的纯相关程度。以三个变量之间的关系为例，如果自第一个和第二个变量中把第三个变量的共同解释力(影响力)排除后，再检视第一个和第二个变量之间相关($r_{12.3}$)程度的变化，通常原本的系数值会变小，或者可能转变成为零相关。净相关的计算公式为：

$$r_{12.3}=\frac{r_{12}-r_{13}r_{23}}{\sqrt{1-r_{13}^2}\sqrt{1-r_{23}^2}}$$

附注：有关公式中任何两个变量之间相关系数的计算方法，请参阅“积差相关”的说明。

算式所得的自变量与因变量之间净相关程度最高者，将被优先选入分析，这也代表了此自变量对因变量的解释力、影响力最高。

部分相关则是指第一个变量与排除第三个变量解释力后的第二个变量之间的相关情形，换言之，第一个变量中仍然含有第三个变量对它的影响成分。例如，自语文成绩(X_2)中排除了智力(X_3)的影响力，但是数学成绩(X_1)中依然包含了智力(X_3)对它的影响作用。因此，在此种情形之下探讨数学与语文成绩的关联性，并无法否定数学成绩仍然还是会受到智力的影响。其计算公式为：

$$r_{1(2\cdot 3)}=\frac{r_{12}-r_{13}r_{23}}{\sqrt{1-r_{23}^2}}$$

(4) 多元相关系数

$$R_{Y\cdot 1234\cdots p}=\sqrt{\beta_1 r_{Y1}+\beta_2 r_{Y2}+\cdots+\beta_P r_{YP}}$$

其中 $\beta=b\frac{S_X}{S_Y}=b\frac{\sqrt{\frac{\sum(X-\overline{X})^2}{N}}}{\sqrt{\frac{\sum(Y-\overline{Y})^2}{N}}}$

R^2 为“决定系数”，数值大小代表预测变量 X 对效标变量 Y 所能解释、影响的程度(通常以百分比表示)。事实上，在多元回归分析中，每增加一个预测变量(X)都会使得 R^2 增大，因此通常需要将公式调整如下：

$$\hat{R}^2=1(1-\hat{R}^2)\frac{N-1}{N-K-1}$$

(5) 总变异 SS_t(total variance)、残差平方和 SS_{res}(又称未解释变异 unexplained variance)、回归离差平方和 SS_{reg}(又称已解释变异 explained variance)

$$SS_t = SS_{reg} + SS_{res}$$

$$R^2 = \frac{SS_{reg}}{SS_t} = \frac{\left[\sum XY - \frac{\sum X \sum Y}{N}\right]^2}{\left[\sum \chi^2 - \frac{(\sum X)^2}{N}\right]\left[\sum Y^2 - \frac{(\sum Y)^2}{N}\right]}$$

因为 $SS_{res} = SS_t - SS_{reg} = SS_t - (R^2 SS_t)$

所以 $SS_{res} = (1 - R^2) SS_t$

(6) 检定所增加的 R^2 显著性(多元相关系数显著性考验)

$$F = \frac{R^2 / P}{(1 - R^2)/(N - P - 1)}$$

其中 P 代表预测变量(自变量)的数目,N 代表总人数,自由度为 $(P, N-P-1)$。

3. 范例

某研究者欲探讨哪些因素会影响社工人员的职业生涯抉择程度。经调查与单因子变异数分析后,将“择业参考因素、个人处事计划程度、工作领域、职位”等 13 个预测变量纳入多元逐步回归检测,资料处理结果摘要如表 13-10 所示。从表 13-10 所列资料可以了解,社工人员择业参考因素中“自我价值”(X_1)的解释力最高(38.92%),即此因素对当事者职业生涯抉择程度的影响力最大,加入预测变量“管理规划度”(X_2)后,对社工人员职业生涯抉择程度的解释力增加了7.21%,而纳入第三个预测变量“社工人员处事计划程度”(X_3)后,又增加了 2.65%的解释力,再加入受访者“职务”(身为机构主管、督导与否,X_4)因素,则可增加 1.91%的解释力,加入第五个预测变量“自我条件”(X_5)后,再增加了 1.12%的解释力,最后,加入“择业参照因素的外控程度”(X_6),则略增了 0.97%的解释力。整体而言,六个预测变量对效标变量的解释(影响)程度为 52.78%。而依据逐步回归分析结果,社工人员职业生涯抉择程度(Y)的最佳预测方程式为:

$$Y = 22.17 + 1.33X_1 + 0.92X_2 + 0.33X_3 + 1.47X_4 + 0.58X_5 - 0.38X_6$$

表 13－10　社会工作人员职业生涯抉择的逐步回归分析

逐步自变量	净回归系数	标准误	标准化净回归系数	净 F 值	R square 及总 F 值	R square 增加量
1. 常数	30.618 6	1.285 5		567.334***	R^2＝0.389 2	
自我价值（择业参考因素）	2.296 5	0.134 6	0.623 8	291.143***	F＝291.143***	
2. 常数	26.833 5	1.302 0		424.784***	R^2＝0.461 3	0.072 1
自我价值	1.884 7	0.137 1	0.512 0	189.080***	F＝144.464***	
管理规划度（择业参考因素）	0.975 4	0.124 8	0.291 0	61.074***		
3. 常数	20.926 1	1.795 4		141.462***	R^2＝0.487 8	0.026 5
自我价值	1.631 3	0.143 6	0.443 1	129.042***	F＝144.464***	
管理规划度	0.957 7	0.121 9	0.285 7	61.740***		
处事计划程度	0.360 4	0.074 2	0.177 7	23.573***		
4. 常数	21.596 4	1.735 6		154.836***	R^2＝0.506 9	0.019 1
自我价值	1.588 8	0.141 4	0.431 6	126.216***	F＝161.691***	
管理规划度	0.923 6	0.120 0	0.275 5	59.240***		
处事计划程度	0.343 0	0.073 0	0.169 1	22.058***		
主管、督导（职称）	1.698 5	0.405 1	0.140 0	17.581***		
5. 常数	19.536 2	1.831 1		113.828***	R^2＝0.518 1	0.011 2
自我价值	1.291 7	0.167 2	0.350 9	59.664***	F＝97.423***	
管理规划度	0.887 0	0.119 3	0.264 6	55.290***		
处事计划程度	0.326 9	0.072 4	0.161 2	20.364***		
主管、督导（职称）	1.658 4	0.401 1	0.136 7	17.096***		
自我条件（择业参考因素）	0.587 9	0.181 1	0.138 7	10.540***		
6. 常数	22.166 4	2.011 0		121.511***	R^2＝0.527 8	0.009 7
自我价值	1.325 2	0.166 1	0.360 0	63.656***	F＝84.194***	
管理规划度	0.920 9	0.118 8	0.274 7	60.137***		
处事计划程度	0.328 2	0.071 8	0.161 8	20.893***		
主管、督导（职称）	1.470 7	0.402 3	0.121 2	13.365***		
自我条件	0.576 4	0.179 5	0.135 9	10.309**		
外控程度（择业参考因素）	－0.379 8	0.125 1	－0.100 1	9.215**		

* $p<0.05$　　** $p<0.01$　　*** $p<0.001$

虚拟变量基准组：教育程度（大学毕业）、宗教信仰（天主教及其他）、工作领域（妇女、家庭及其他）、职称（其他）、聘用情形（其他）

承上所言,在社工人员的职业生涯发展过程中,若其重视“自我价值”、其工作部门的“管理规划”明确、“个人处事计划程度”高、身兼“部门主管或督导职务”、拥有从事社工的“自我条件”,以及较不重视工作场域的“外控因素”(如交通、人际、亲友支持与否)等条件的,其选择投入社会工作行业,而且以此行业作为终身职业的认定程度便愈高。

其他并入分析的变量,如宗教信仰、工作领域、聘用情形、职业生涯规划策略导向、薪酬制度等预测因素,对职业生涯抉择(效标变量)总变异量的解释力未达统计显著水准,因此可以略过因果关联的考量。

(二)阶层式复回归分析

阶层式回归分析以理论及实证逻辑原则来决定各个自变量进入回归分析模型的先后顺序。换言之,研究者主控了各个自变量在不同阶段中被纳入检视的顺位。此外,阶层式回归与逐步回归的统计学原理及计算公式极为相似。

1. 适用时机

(1)建构理论(依据变量之间的因果逻辑及发生时间的前后顺序)。

(2)变量测量性质与逐步回归分析相同。

2. 公式

使用的运算元素及方法与逐步回归分析相同。

3. 范例

医院社工人员欲关怀慢性病人照顾者的健康情形,于是对相关对象进行了“健康状况影响因素”的问卷调查。研究者依据研究假设将各个预测变量分阶段纳入讨论,因此自变量被归入分析的先后顺序是由研究者所决定和掌控的。表 13－11 所列的数据,即是经整理后的研究发现。由于照顾者本身所拥有的个人特质如年龄、性别、教育程度、婚姻、工作、居住状况等因素乃是影响个人健康状况的先决条件,因此它们最早(第一阶层)被引入分析。表中资料显示,“照顾者个人特质”对其整体健康状况的解释力达 11.6%(可以解释照顾者健康状况总变异量的 11.6%);其次,纳入患者病前与照顾者之间的“情感关系”后,所增加的解释力并不明显;第三阶层纳入“病人医疗特质”(罹病系统、就医保险身份等)加以检测,则可增加 3.9%的总变异解释量;第四阶层纳入“病人自我照顾能力”,则增加了 2.8%的变异解释量;第五阶层将“照顾者负荷”(有无分担人手、可以分担照顾工作的人手数)并入考量,则又可增加 3.7%的预测解释力;最

后纳入中介变量“压力知觉程度”再予分析，其对“照顾者健康状况”变异量的解释力剧增了 33.7%。综合言之，六个自变量群组对“照顾者健康状况”产生了 55.8%的解释力，其中又以最后一阶层“压力知觉程度”的影响力最高。

表 13－11　照顾者健康状况层级回归分析

变　量　组	回归系数	R^2	增加的 R^2	p 值
照顾者基本特质	0.341	0.116		0.019*
情感关系	0.341	0.116	0.000	0.762
病人医疗特质	0.395	0.156	0.039	0.017*
病人自我照顾能力	0.429	0.184	0.028	0.049*
照顾负荷	0.470	0.221	0.037	0.044*
压力知觉程度	0.747	0.558	0.337	0.000***

* $p<0.05$　　*** $p<0.001$

层级回归分析的结果，说明了照顾者自觉的生理、心理压力程度以及病人罹病系统（是否罹患癌症）均显著影响着照顾者的健康状况（参见表 13－12）。换言之，照顾者生理、心理压力愈大，而且照顾的对象是癌友时，则其健康状况将愈差。

表 13－12　照顾者健康状况影响因素的层级回归系数分析

变　量	1		2		3		4		5		6	
	b	Beta	b	Beta	b	Beta	b	Beta	b	Beta	b	Beta
居住状况	2.104	0.072	2.075	0.071	2.316	0.079	1.708	0.058	1.598	0.055	2.536	0.087
婚姻	1.206	0.042	1.234	0.043	0.411	0.014	0.451	0.016	0.695	0.024	−0.337	−0.012
性别	−1.480	−0.056	−1.409	−0.053	−3.447	−0.013	−1.106	−0.042	−1.228	−0.046	−1.012	−0.068
教育程度	−0.853	0.063	−0.886	−0.065	−0.852	−0.063	−1.128	−0.083	−0.690	−0.051	0.040	0.003
工作状况（有、无）	2.593	0.120	2.628	0.121	2.893	0.133	2.265	0.104	1.891	0.007	1.625	−0.075
年龄	−0.036	−0.052	−0.038	−0.054	−0.077	−0.111	−0.065	−0.094	−0.075	−0.107	−0.073	−0.104
关系（配偶非配偶）	−6.312*	−0.304	−6.195*	−0.296	−4.816	−0.230	−5.511	−0.263	−4.437	−0.212	−1.206	−0.058
情感关系			−0.501	−0.022	−0.606	−0.028	−0.144	−0.051	−0.771	−0.034	−0.001	−0.000
癌症与否					−2.169	−0.106	−2.949*	−0.144	−3.829*	−0.187	−3.217**	−0.157
身份					4.635**	0.193	4.358*	0.181	4.405*	0.183	2.574	0.107

(续　表)

变　量	1		2		3		4		5		6	
	b	Beta	b	Beta	b	Beta	b	Beta	b	Beta	b	Beta
基本生活功能							0.072	0.036	−0.019	−0.009	−0.120	−0.060
辅助性生活功能							0.371	0.154	0.451	0.188	0.236	0.098
信息支持									0.959	0.039	−1.293	−0.053
有无其他照顾人手									−4.465*	−0.148	−1.879	−0.062
照顾人手数									1.296*	0.191	0.430	0.063
生理压力											−1.862**	−0.498
心理压力											−0.627*	−0.171
社会压力											−0.145	−0.033
常数	48.278		48.423		48.854		46.207		45.865		81.284	

* $p<0.05$　** $p<0.01$　*** $p<0.001$, $N=190$

承前(*p*.457),其他被纳入分析的预测变量如照顾者是否与病人同住、照顾者婚姻状态、两者病前情感关系等因素,并未对照顾者的健康形成直接而显著的影响,其可能借由照顾者的压力知觉而对健康形成间接影响。据此,照顾者健康状况的最佳预测模型,可以下列方程式表示:

$$\begin{array}{c}\text{照顾者健}\\\text{康状况}\end{array}=81.28-\left(1.86\times\begin{array}{c}\text{生理压力}\\\text{知觉分数}\end{array}\right)-\left(0.63\times\begin{array}{c}\text{心理压力}\\\text{知觉分数}\end{array}\right)-\left(3.22\times\begin{array}{c}\text{病人罹}\\\text{患癌症}\end{array}\right)$$

第四节　实务研究与资料分析的其他议题讨论

电脑软件的开发,使得较为繁复的统计技术诸如路径分析(path analysis)、区别函数分析(discriminant function analysis)、复回归分析(multiple regression)以及逻辑回归(logistic regression)等问题,均得以借助软件的高速运算而获得解决。以下即针对实务研究处理量化资料较常遇到的几项问题予

以说明。

一、问卷设计的信度与效度分析

社会工作实务研究经常使用自行设计的问卷作为了解行为现象之间关系的媒介。因此针对问卷信度与效度的讨论，也就成为介绍研究设计时必须说明的项目。目前经常被实务研究使用的信度、效度分析方法包括：建构效度（construct validity）——以因素分析（factor analysis）模式解释；内部一致性分析（analysis of internal consistency）——以 Cronbach's alpha 系数证明题目设计的一致性或稳定性。此外，为了检视问卷试测（pilot study）后的适当性，项目分析（item analysis）问卷题目的关联性分析（correlation analysis）以及鉴别指数（discrimination index）等，都是研究者较常采用的信度评量工具。

（一）建构效度的因素分析法

1. 概念释义

通过变量内在成分之间的关联性分析，将原始变量所组成的内涵予以精简化，抽选数个潜在性因子结构（共同性）予以命名，并由这些因子形成新的线性组合，来代表该变量或概念的意涵。简言之，因素分析法的功能是协助研究者寻找最具代表性的可观察项（变量的核心成分），并以其表示某个抽象的不可观察项或概念。

2. 分析步骤

（1）计算相关矩阵

项目之间的相关属于 Q 式因素分析法。通常因子与项目之间的相关系数最好达到 0.30 以上，即其因素负荷量（factor loading，代表“项目”与“各项目组合而成的因素”之间关系强度的数值）超过 0.30 的标准方予选取。

（2）抽取最初的共同因素

以主成分分析法（principal component analysis），在正交（将各因素之间视为独立或不重叠的情况）成分中抽取最大变异量因素。即以最高相关系数法，求得各共同因素的特征值或固定值（eigen-value，指因素矩阵中每列因素负荷量的平方和，代表各因素或观察值对于各变量变异量解释力的贡献程度）。一般而言，保留特征值大于 1 的共同因素。

(3) 进行最后的因素转轴

当因素之间被视为彼此独立时,多采用正交(直交)转轴法(orthogonal rotation)使各因素所能解释的变异量趋向最大(varimax)。社会科学所观察的行为现象,多半是相互影响而彼此牵动的状况,因此,斜交转轴法(oblique rotation)才是较为适合的选择。换言之,多数的变量及概念或多或少都蕴含了其他变量、概念的因素成分。例如,"社会声望"与"社会经济地位"两个概念都涵盖了教育程度、职业、收入等组合成分,此时就需要采用斜交转轴法,将重叠的成分予以归并处理,才能够促使观察项达到最佳代表性。

(二) Cronbach's alpha 系数信度分析法

1. 概念释义

Cronbach's alpha 系数信度分析法适用于态度量表(多半为五点计分方式设计)的分析。它是一种用以评估量表或问卷一致性(consistency)或稳定性(stability)的指标。

2. 公式算法

$$\alpha=\left(\frac{K}{K-1}\right)\left[\frac{S_x^2-\sum S_i^2}{S_x^2}\right]$$

式中,K 为测验的总题数;S_x^2 为测验总分的变异数;S_i^2 为每个题目分数的变异数。

3. 分析步骤

(1) 先进行量表题目的筛选,去除与总分之间相关系数较低的题目。

(2) 继续运算总量表与分量表(变量所包含的向度)之间的信度系数。

(三) 问卷试测后的关联性考验与鉴别指数分析

1. 概念释义

(1) 关联性分析(correlation analysis)

通过每一题目与总分之间的积差相关系数,来评估受试者态度反应的一致性或稳定性。一般而言,当相关系数低于 0.30,且未达到相关性显著考验水准($\alpha=0.05$)的题目将予以删除。

(2) 鉴别指数(discrimination index)

通过每一题的鉴别力(discriminatory power)分析,来判断题目设计是否能

够分辨受试者的态度反应。计算鉴别力的大小，就是将所有受试者问卷测试的总分依照高低顺序排列，再将27％分数居前者（或以25％为标准）列为高分组，27％分数居后者（或以25％为标准）列为低分组，最后计算高低分组在各个题目平均数的差异值，以其代表鉴别力；并借由决断值（critical ratio）显著性检定决定该题目的保留、修订或删除。

2. 公式算法

（1）积差相关系数的显著性检定

以往多利用统计学专业书中的附表直接检视 r 系数的显著性。目前则可由电脑软件以下列公式计算后加以判定。

$$t=\frac{r-\rho}{\sqrt{\frac{1-r^2}{N-2}}}$$

ρ 为总体相关系数，通常设 $\rho=0$；$\mathrm{d}f$（自由度）$=N-2$

（2）鉴别力分析

$$DP=\overline{X}_{\mathrm{H}}-\overline{X}_{\mathrm{L}}$$

式中　$\overline{X}_{\mathrm{H}}$ 为高分组在某题得分的平均数；$\overline{X}_{\mathrm{L}}$ 为低分组在某题得分的平均数。

$$CR=\frac{\overline{X}_{\mathrm{H}}-\overline{X}_{\mathrm{L}}}{\sqrt{\frac{S_{\mathrm{H}}^2}{N_{\mathrm{H}}}}+\sqrt{\frac{S_{\mathrm{L}}^2}{N_{\mathrm{L}}}}}$$

式中　S_{H}^2 为高分组在某题得分的变异数；S_{L}^2 为低分组在某题得分的变异数；N_{H} 为高分组人数；N_{L} 为低分组人数。

$$\mathrm{d}f=N_{\mathrm{H}}+N_{\mathrm{L}}-2$$

二、虚拟变量(dummy variable)的处理

（一）概念释义

此为资料编码处理的方式，通常是将类别或顺序尺度的自变量观察资料（如性别、宗教信仰、受教育程度等）转换成为等距、等比尺度的数值，如此便可投入多元回归预测方程式中予以分析讨论。

(二) 处理方式

将原类别(或顺序)变量各个类属都视为一个变量(虚拟),每一个观察值对此类别而言只有是(以 1 代表)或不是(以 0 代表)的两种可能性。操作时,k 个类别的变量只须转换成$(k-1)$个虚拟变量,其中被排除在外的类别将作为解释虚拟变量效果的参考基准组。例如,受试者婚姻状况(属性变量)区分为已婚、未婚、其他三个类别,故可产生两个虚拟变量,以表 13-13 说明。

表 13-13 婚姻状况的虚拟变量处理

每一虚拟变量的分数			效果说明
虚拟变量名称			期望值(Y')
婚姻状况	D_1	D_2	
已婚	1	0	$a+b_{11}$
未婚	0	1	$a+b_{12}$
其他	0	0	a

回归方程式:$Y'=a+b_{11}X_{11}+b_{12}X_{12}$

此为原始方程式:$Y=a+b_1X_1$。婚姻状态已由 X_1改成而 X_{11}(已婚)、X_{12}(未婚)与 X_{13}(其他)三个变量,而其中 $b_{13}X_{13}$因被视为基准组,所以等于 0,无须列入方程式。

三、误失(遗漏)值(missing value)的处理

(一) 概念释义

指受访者对问卷中的某一或某些题目漏答或拒绝回答,因而造成资料空白的情况。当样本人数不多而收集到的资料又出现许多遗漏时,该项研究成果与结论将会受到严重质疑。

(二) 处理方式

(1) 通常是以空白或以某些不可能出现在题项中的数值作为替代,并予以编码,如以 0 或 9 代表漏填值。

(2) 以该变量或该题的平均值(mean)替代(依据趋中原理,如果同一研究频繁使用均数替代法,则相关系数可能会有被低估的风险)。

分析此类资料大致上包括两种做法。其一是将失误值不列入运算,但此种

做法必须先就该题目的重要性予以评估，此外尚须考量问卷中是否有其他题目足以替代。其二则被视为较为妥当的处理方式，即先将该题目或该变量的回答者与未回答者分为两个组别，而后再比较这两组受试者在其他题目或变量上的态度反应是否也有差别。如此分析，有助于研究者再行检视问卷设计的题目陈义与用词是否周全明确。

四、研究样本数的选取标准

估算样本数目是研究计划阶段就必须决定的问题。适当的样本数不但可以获得有价值的研究成果，而且能够节约成本。决定样本数目通常会受下列因素的影响：(1) 研究者的时间、人力与财力状况；(2) 研究性质(量化或质性研究)与目标(预定达到的研究目的)；(3) 研究群体的同质性程度。

以下几项样本数目选取标准，供研究者参考。

(一) 信赖区间作为取决标准

以样本平均数来估计总体平均数的信赖区间、水准，进而估算研究所须抽取的样本数目。即研究者已经决定了样本与总体平均数的误差值(e)及显著水准(α)。计算公式包含两种情况。

1. 已知总体标准差 σ

$$N=\left[\frac{Z_{(1-\alpha/2)}\sigma}{e}\right]^2$$

(Z 分数依据显著水准判定，属双尾检定)

因为$\overline{X}-Z\left(\frac{\sigma}{\sqrt{N}}\right)\leqslant\mu\leqslant\overline{X}+Z\left(\frac{\sigma}{\sqrt{N}}\right)$

而$|\mu-\overline{X}|=e$(误差值)

所以 $e=|\mu-\overline{X}|=Z\left(\frac{\sigma}{\sqrt{n}}\right)$

两边平方后得 $N=\left(\frac{Z\sigma}{e}\right)^2$

2. 未知母体标准差 σ

$$N=\left[\frac{t_{(1-\alpha/2,\ N-1)}\hat{S}}{e}\right]^2$$

$\hat{S}=\sqrt{\frac{\sum(X-\overline{X})^2}{N-1}}$,此为总体标准差的不偏估计值 t 分数之 $df=N-1$,依显著水准判定,属双尾检定。

因为 $\overline{X}-t\left(\frac{\hat{S}}{\sqrt{N}}\right)\leqslant\mu\leqslant\overline{X}+t\left(\frac{\hat{S}}{\sqrt{N}}\right)$,同理得 $N=\left(\frac{t\hat{S}}{e}\right)^2$。

此种取决标准的估算通常会面临两项难题,其一是如何制订合理的误差值;其二是研究包含多个变量时,须先判定何者为关键变量,而后再依据关键变量的情况决定误差容忍度。

3. 总体为二项分配的状况

$$N=\frac{Z^2_{(1-\alpha/2)}}{4e^2}$$

因为 $\hat{P}-Z\sqrt{\frac{P(1-P)}{N}}\leqslant P\leqslant\hat{P}+Z\sqrt{\frac{P(1-P)}{N}}$

其中,$\hat{P}=\frac{f}{N}$(样本百分率),而 $E(\hat{P})=P$(总体百分率)。

而 $|\hat{P}-P|=e=Z\sqrt{\frac{P(1-P)}{N}}$

其中,P 未知,常以极大值 $\frac{1}{2}$ 估计,使 $P(1-P)$ 的乘积为最大(得到 N 值的最高临界值,upper bound),故将 $P(1-P)=\frac{1}{4}$,代入算式中,得 $N=\frac{Z^2_{(1-\alpha/2)}}{4e^2}$。

(二)抽样理论作为取决标准

探讨变量关系的研究至少要有 30 个样本,而比较性研究与实验性研究每组至少要有 15 个样本。调查性研究每组至少要有 100 个、次要组别则至少要有 20～30 个样本(欧沧和、李茂能,1992)。

(三)统计方法作为取决标准

1. 多元复回归分析

统计学家 Tabachnick 及 Fidell(1983)建议,多元复回归分析的样本数目最好是所有分析变量数目的 20 倍以上,最少则不要低于变量数目的 4～5 倍(王丛桂、罗国英,1992)。

2. 因素分析

当受试者同质性高且研究变量数目较少时，样本数在 100～200 已足够。此外 Comery（1973）的建议标准则是：50 人——很差，100 人——差，200 人——尚可，300 人——良好，500 人——很好，1 000 人——极佳（李金泉，1993）。

关键名词

t 检定（t-test）：用以考验两个小样本（$N \leqslant 30$；$N=50$）之平均数是否具有显著差异的统计考验方法。

Z 分数（Z score）：利用直线转换的原理，将原始分数与其平均数之差作为分子，以其标准差作为分母所形成的一种没有单位的相对地位量数。

已解释变异量（explain variance）：指一个效标变量的总变异量中，能由一个或一组预测变量加以解释的百分比例。

不对称关系（asymmetrical relationship）：自、因变量之间呈现因果关联性。即自变量会影响因变量而因变量不会影响自变量的关联模式。简言之，两者呈现因果关系。

内部一致性分析（analysis of internal consistency）：用于表示试题或问卷题目鉴别度的分析方法。目的是为检查个别试题（题目）与整体测验（问卷）之间的设计内涵是否稳定一致。其计算方法包含两种情形。

1. 鉴别指数（discrimination index）：比较高分组和低分组在个别试题（题目）上答对人数百分比或选项得分的平均数是否具有清楚的区别性。即将全部试测样本依总分高低排列，取最高及最低的 27%（或 25%）作为高、低分组，再以 $D=P_H-P_L$ 求其鉴别指数 D。

2. 关联性分析（correlation analysis）：求个别试题（题目）与测验（问卷）总分之间的关联性，即以在该题上的对错为二分变量，测验总分为连续变量，求二系列相关（γ_{bis}）或点值二系列相关（γ_{pb}）。

此外，依据 Likert scale 原理所设计的态度量表，虽属顺序尺度的测量，但在资料统整技术上，多半将其视为分数（连续变量）加以处理，故宜以 Pearson's 积差相关系数（r）方法进行分析。

卡方考验(chi square test)：一种无母数统计显著性检定法。通常用于处理类别及顺序尺度的资料。

未解释变异量(unexplained variance)：指效标变量(因变量)的总变异量(变化程度)中，不能由研究设计的预测变量(自变量)所解释的变异量百分比例。

母数或参数(parameter)：代表总体性质的量数。

母数与无母数统计学(parametric and nonparametric statistics)：应用于假设检定的推论统计方法中。前者对于总体资料订有基本假设，包括资料在等距尺度以上的测量层次、样本来自常态分配的总体，以及总体的变异数相等等条件。主要的母数统计方法有 Z 考验、t 考验、变异数分析、积差相关及多元回归分析等。后者对于总体资料分配没有要求特殊的条件，此类统计方法并不涉及总体母数(如 μ 或 σ 的推估)，因此统计检定力比母数统计方法薄弱。常使用的无母数统计方法有 χ^2 的适合度考验、连检定、中数考验及符号检定等。

共变数分析(analysis of covariance)：利用控制方式(例如将实验的各组平均数予以调整)将可能影响实验正确性的误差加以排除的统计方法。

因素分析(factor analysis)：协助研究者找出最具代表性的可观察项(变量的核心成分)，并以其来表示某个抽象的不可观察项或概念的统计方法。

列联相关系数(contingency coefficient)：一种用于测量两个类别变量相关程度的统计方法。

事后比较(a posteriori comparisons)：观察两组以上受试者的实验(处遇)效果时，先以变异数分析组群平均数差异，当变异数分析的 F 值达到显著水准后，再进一步考验$\frac{k(k-1)}{2}$组别中哪几对的平均数差异具有显著性，此种检定方法即为事后比较。

重复量数(repeated measures)：对相同的研究对象在不同时间内予以测量，或者对相同研究对象施予不同的实验处遇后(如不同的辅导方案)再行测量，以比较其中差异变化的情形。

效标变量(criterion)：在回归分析中，指被一个(简单回归)或多个(复回归)预测变量所解释或预估的变量，在一般实验研究设计里，即为受到自变量(因)影响的因变量(果)。例如，以学历来预测收入状况，“学历”为预测变量，而“收入”则是效标变量。

消减误差比例(proportionate reduction in error, PRE)：指以 X 值(自变量)来预测 Y 值(因变量)所减少的误差程度与不知道 X 值来预测 Y 值所产生全部误差程度的百分比值。所消减的误差比例愈高，表示 X 与 Y 变量之间的相关程度愈强。

推论统计学(inferential statistics)：以样本资料来推估总体特征与性质的统计方法。根据概率理论，推论统计学包括两项主题：母体的母数估计及假设检定。

统计分析(statistical analysis)：应用统计学的描述或推论统计方法，将原始资料所呈现的庞杂数据予以简化、归整或推估计算的过程。

统计量：任何由样本观察值(数据)构成的函数均可称为统计量。例如，算术平均数 f 可描述观察值的集中趋势，标准差 S 则可描述观察值的分散或差异趋势。

逐步回归(stepwise regression analysis)：它属于多变量分析的统计方法。将自变量逐步(一次一个，或一次一群)代入回归方式中，依序分析各个自变量对因变量变异的解释力或影响程度。资料处理时，自变量纳入分析的顺序是依照自、因变量之间净相关系数大小来决定的，系数愈高，愈早被选入分析。

区别分析(discriminant analysis)：以多个等距或等比变量来区辨两个或两个以上组别的异同性质。其区别系数已经过标准化转换，因此可以直接代表预测变量对效标变量(区别函数)的解释力或影响程度。

连续变量(continuous variable)：变量的单位可以不断细分的情况属连续变量。例如，身高、体重、年龄、时间、音量、速度……通常等距或比率(等比)尺度所测得的变量属于连续变量。

通则取向(nomothetic approach)：将研究概念、变量转化成为可以测量、定义、操作的数值，经由系统化与标准化方法收集资料，再运用、分析大量样本，以借之建立社会现象或事件之间的一般性通则或法则。

单尾(侧)或双尾(侧)考验(one-tailed test or two-tailed test)：进行假设考验时，凡检定单一方向性的问题，如“大于”“高于”“优于”等情形，谓之单尾考验。当检定的重点置于两总体平均数有无差异，或是两变量之间是否具有关联性时，如“等于”“不等于”“相关”“不相关”则为双尾考验。后者因较难达到显著水准(须除以 2)，因此比较不容易拒绝虚无假设，而其触犯型误差(拒绝虚无假设

所犯的错误)的概率也相对较小。

描述统计学(descriptive statistics):以计划、归类、计算等方法,将大量资料加以整理、归整、摘要,以了解资料的形态、特质及其内部关系的统计学类。

虚无假设(null hypothesis):变量间的关系状态与研究者心中所欲支持的论点刚好相反的假说即为虚无假设。当实证资料否证了它的真实性时,则表示间接证实了研究者的论点可能为真。

虚拟变量(dummy variable):将类别变量中的每一类别指派为 0 或 1(否或是),目的是让类别尺度的资料亦能进行回归分析的一种资料转换处理方法。

量化资料(quantitative data):以数量(值)呈现其特质的如社会现象等资料。

测量层次(level of measurement):变量的数值或属性具有不同的运算性质。分析资料必须依据变量属性而选择适当的方法。一般量表或变量设计的测量尺度,运算性质由低至高共分为四个层次:类别、顺序、等距与等比(比率)尺度。

阶层式复回归分析(hierarchical multiple regression analysis):它在统计学原理上同于逐步回归分析,唯其自变量纳入复回归模型分析的顺序,由理论上因果关系的先后逻辑来决定。

项目分析(item analysis):检验量表或问卷编制的题目在难易度、鉴别力上的适切性。

间断变量(discrete variable):变量的单位不可再细分,一定为整数的情形。例如,人数、类别数、件数,通常类别或顺序尺度所测得的变量属于间断变量。

解析性研究(analytical research):以探讨变量间关系的因果顺序与因果模型为目的的研究类型。

路径分析(path analysis):运用一系列多元回归分析及解释各变量间(通常具有时间逻辑顺序)线性关系的方法。研究者须依据理论及经验决定相关变量的结构模式。通常因果关系建立的次序为单向模型:某些变量可能引发其他变量的变化,但其后发生的变量绝不可能有反馈影响的情形。

预测变量(predicator variable):借着回归方程式预测某一效标变量的数值时被研究者所操作的变量(如服务方案、家庭支持程度)即为预测变量。例如,$Y=b_1X_1+b_2X_2+a$,此方程式中,Y 为效标变量(例如社会适应状况),而 X_1 与

X_2 则是预测变量。另外，在因果关系模型中，效标变量通常受到研究者所操作的自变量（因）的影响，而被用来分析自变量（因）对因变量（果）的影响程度。

对立假设（alternative hypothesis）：研究者所支持的变量间关系状态的陈述假说。通常以两个变量之间具有关联性或两个总体平均数之间具有差异性的论点呈现。

对称关系（symmetrical relationship）：自变量与因变量之间的关联性并不区分或不确定彼此影响的方向或顺序。简言之，两者呈现共变关系。

标准分数（standard score）：说明个人在某变量（行为表现）上相对位置的数量表示法。它是以标准差为基准，说明个人分数距离群体或全体平均数有多少个标准差距单位。通常此分数均设定在常态分布的条件下，Z 分数即是最典型的标准分数。

标准差（standard deviation）：代表样本之间离散或分散情形的量数或指标。

积差相关（product-moment correlation coefficient）：应用 X、Y 变量的共变量与两变量标准差乘积的比值，来代表两个等距或等比尺度变量之间关联程度的统计数。

独特取向（idiographic approach）：对特定主体或对特定对象的叙说、观察、讨论所获得的原始资料，经由“过录”、分类、赋予概念的分析过程，而达到透析与诠释行为现象含义的方法。此一研究形式不追求也不强调社会现象或事件经由化约概化后的通则性。

属性变量（attribute variable）：实验研究中研究者不能操作而只能描述的变量，如性别、居住地区、宗教信仰、教育程度等。属于类别尺度的测量层次，通常被当作自变量处理。

变异数分析（analysis of variance，ANOVA）：用以检定两个以上样本群或总体平均数差异的统计方法。

逻辑回归（logistic regression）：用以分析、解释一个以上之等距、等比尺度的自（预测）变量与类别尺度的依（效标）变量之间关系的统计方法。逻辑回归的基本假设与线性回归类似，主要差别在于因变量属于二元化资料，如成功或失败、有效或无效、低收入家庭或非低收入家庭。

习　题

1. 试说明量化资料分析所应用的统计方法的选判原则包括哪些。
2. 试说明各种双变量关联系数分析的适用条件。
3. 试举例说明各种卡方考验的类型。
4. 当双变量均属等距或等比测量尺度,且彼此之间呈现对称关系,而研究者欲分析变量之间的关联性时,较常使用的统计方法为何?试举例说明之。
5. 试举例说明假设考验的五个步骤。
6. 试分析 t 检定的功能、适用时机,并举例说明其各种检定的类型。
7. 试说明单因子变异数分析的适用时机、检定类型,以及较常使用的事后比较方法的适用条件。
8. 当研究者欲分析行为现象之间的因果关系模型,抑或建构行为现象的最佳预测模型时,试问较常使用的统计方法是什么?各种方法的适用时机又有何差异?
9. 试以任何一篇研究调查报告为例,讨论其如何检测问卷的信度与效度,并请提供相关建议。
10. 何谓虚拟变量?试举一例说明虚拟变量的处理方式。
11. 试说明如何处理研究资料中的误失(遗漏)值。
12. 试问研究样本大小的决定包括哪几方面的参照准则?

第十四章 论文写作

第一节　前言

第二节　研究论文的次序与格式

第三节　论文第三章——“方法论”的写作

第四节　论文第四章——“分析与发现”的写作

第五节　论文第五章——“结论”的写作

第六节　参考文献与附录

第七节　其他

摘　要

研究的进行毕竟是繁重的工作，从研究的构思到资料的阅读，从方法的拟订到问卷的设计，以及资料的收集与分析等内容，每一个步骤都有基本的要求，甚至都有一定的格式。在每个阶段，研究者必须巨细靡遗、深思熟虑地去进行，由不得丝毫的松懈与怠慢。关于如何完成一份研究论文，有关撰写的一些原则、论文内容应包含哪些项目、方法论、资料的收集与分析、最后的结论与建议等的写作，与其他注意事项等，都将在本章做介绍，提供给读者参考。

第一节　前　　言

研究的进行毕竟是繁重的工作，从研究的构思到资料的阅读，从方法的拟订到问卷的设计，从资料的收集到统计的分析，研究者需巨细靡遗、深思熟虑，由不得丝毫的松懈与怠慢。然而，一般人将研究做到资料分析时大概都是“气势将尽”，或是为了能在规定的时间内交卷，常在最重要的论文写作时力不从心，使整个论文的呈现千疮百孔、惨不忍睹、相当可惜。

研究论文的写作并不是小学生第一次的作文，每每要绞尽脑汁才能寸步维艰、字字血泪地完成。研究论文其实很简单，因为它有一定的格式、步骤，只要依其格式，按部就班地把研究内容与过程平实地呈现出来就是合格的研究写作。当然，研究报告不是一般的抒情文章，可以任作者想到哪里就写到哪里，一个好的研究报告应该合乎某些标准，如文字与资料的陈述应该清楚、精确、简明；文章的主题必须前后连贯、有方向可寻；分析与结论应该客观严谨；文法与体裁也应合乎一般的格式（Beach & Alvager，1992），兹简单说明如下。

第二节　研究论文的次序与格式

一、研究论文的次序

依学术惯例，研究论文的章节次序事实上是固定的。

第一章：绪言（introduction）
第二章：方法论（methodology）
第三章：分析与发现（finding）
第四章：讨论、建议与结论（conclusion）

在这种格式中,第一章的“绪言”事实上包括了问题叙述与文献探讨,第二章的“方法论”中则介绍了研究架构、抽样、问卷设计、资料的收集与分析过程等。第三章的“分析与发现”则包含了样本介绍(profile)与各类的资料分析,最后才是讨论、建议与结论。国内目前各校的习惯大同小异,与学术惯例中的四章写法不完全一致,却相距不远,其格式大概如下。

第一章:问题叙述(problem statement)(有的研究以研究宗旨或研究目标开始)

第二章:文献探讨(literature review),文献后则交代名词界定与解释(有的研究把名词界定放在方法论内说明)

第三章:方法论

第四章:分析与发现(包括样本介绍、资料分析等)

第五章:结论[包括研究应用(implication)、讨论(discussion)、建议(suggestion)、研究限制(limitation)等四节]

阅览各式论文报告时,一定会发现论文固然有一定的格式,但是因为学科的不同,在格式上仍会存在一些差异,自然科学的论文格式与社会科学的论文自有所不同,我们先分析一般科学方面的论文格式。

二、一般的科学论文纲要

一般科学论文格式纲要

Preliminary Materials(前置资料)

1. Title page　封面
2. Preface and Acknowledgements　致谢词
3. Table of contents　内容目录
4. List of tables　表格目录
5. List of figures　图形目录

Body of the Paper(论文主体)

Chapter 1. Introduction(序言,导论)

a. General statement of the problem(问题叙述)
b. Statement of the hypotheses(假设)
c. Definition of terms(名词界定)

Chapter 2. Review of Literature(文献探讨)
- a. Review of previous research(先前研究)
- b. Pertinent opinion(特殊意见)
- c. Summery of state of the arts(摘要)

Chapter 3. Methodology(方法论)
- a. Description of subjects/samples(样本叙述)
- b. Description of measures employed(测量方法说明)
- c. Research design and procedures(研究设计与步骤)

Chapter 4. Findings(发现)
- a. Analytic techniques(分析的技术与方法)
- b. Description of findings per hypothesis(有关研究假设的发现)
- c. Other findings(其他发现)

Chapter 5. Summary and Conclusions(摘要与结论)
- a. Summary of hypotheses, method, and findings(假设摘要,方法论摘要,发现摘要)
- b. Conclusions(结论)
- c. Implications(应用)

Reference Materials(参考资料)
1. Bibliography(书目)
2. Index(索引)

三、社会科学论文纲要

以上所呈现的是一般学术界常用的格式,社会工作的论文格式也是大同小异,但若要详细说明,请见下列纲要。

社会科学论文格式纲要

论文前置资料(preliminary materials)
1. 封面(cover)
2. 致谢词(acknowledgement)
3. 论文章节目录表(table of content)
4. 表格目录表(list of tables)
5. 图形目录表(list of figures)

论文本文(body of research)

第一章:前言(introduction)
1. 问题叙述(problem statement)
2. 研究宗旨或研究目标(goal of research)

第二章:文献探讨(literature review)
1. 理论背景(theoretical background)

2. 先前研究(previous study)
3. 相关变量的研究结果(variables of research)

第三章:方法论(methodology)

1. 研究的重点简述(introduction of methodology)
2. 研究架构图(frame of research)
3. 研究设计简介(research design)
4. 研究假设说明(hypotheses)
5. 重要变量的操作定义(operational definition)
6. 量表制作过程说明(measurement)
7. 抽样过程说明(sampling process)
8. 资料收集过程及资料分析策略(data collection and analysis statement)

第四章:分析与发现(analysis and findings)

1. 样本简介(profiles)
2. 分析与发现(analysis and findings)

第五章:结论与建议(conclusion)

1. 结论(conclusion)
2. 应用与建议(implication and suggestion)
3. 研究限制(research limitation)

附录(appendix):

1. 书目(bibliography)
2. 问卷表(questionnaire)
3. 其他相关附件(research documents)

第三节　论文第三章——“方法论”的写作

问题叙述与文献探讨的写作都已在本书前几章中介绍,本处不再重提。方法论的写作是展现研究实力的重要单元,研究者不可忽略。方法论的说明可以使读者很快地了解研究的主题与重点,也可以洞察执行的过程是否正确无误。方法论亦可让读者知晓研究者是否有研究的创造性、突破性与实用性。内行的研究人员,亦可以从研究者的方法论中知晓研究者是否具备研究的能力,洞察他的研究态度与学术态度,因此对于此部分的写作丝毫大意不得。方法论主要有下列重点。

一、重述研究的重点

第二章的文献虽已把研究的概念做了介绍,但是进入研究法这个段落时,

则不能毫无交代地写下去，多少总要把研究的主要概念、研究的方向做个摘要式的说明，使读者脑海里能有更清楚的背景。例如：

“父母亲的婚姻满足程度如何影响了子女的婚姻品质是本研究的重点。本研究一方面必须测量父母亲婚姻的满足程度，另一方面必须探讨子女的婚姻品质。当然，婚姻品质的项目很多，本研究必须详加规划，影响婚姻品质的因素也不只是父母亲的婚姻满足度而已。一些可能的因素，包括子女的属性变量如职业、受教育程度、收入、人格成熟度、婚姻压力指数等都会影响本身的婚姻品质，本研究均将设法探讨。此外，也借着严格的随机抽样与足够的样本数，使一些外加变量被控制，或使其彼此综合……”

当简单说明这些重点以后，研究者再着手把研究方法重点一一陈述，读者就不会觉得太突兀。

二、研究架构图

事实上研究论文中是否需要画出研究架构图并没有定论，随指导教授的偏好而定。然而，不可否认，架构图能使阅读者在很短的时间内对整个研究有通盘的了解，这是不争的事实。如图 14－1 所示。

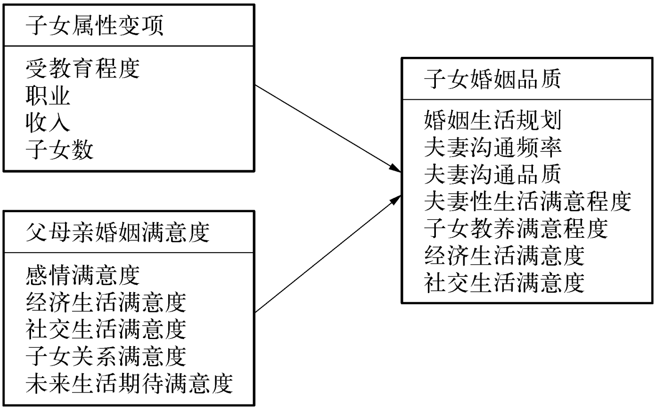


图 14－1　“父母亲婚姻满足程度如何影响子女婚姻品质”研究架构

在架构图中，最重要的单元是自变量、因变量、基本属性与其他相关的变量等。图中，方格与方格之间的箭头如何画是很大的学问，很多人以为箭头代表

两个变量之间的因果关系,由一方指向另一方,代表着第一变量是因,第二变量是果。若是线条的双方都有箭头,表示这两个变量是彼此影响,它们彼此有关,只是不知道谁是因,谁是果而已。然而在社会科学的研究中,实际上因果关系的变量并不多,两个变量的因果关系,其实只是“关联”而已,而不是因果。所以在架构图中所绘制的箭头,不妨把它当作研究者在统计应用时的使用次序。所以在计算自变量如何影响因变量时,这两个变量虽可能是因果关系,但更可能只是彼此关联而已。

三、研究设计的介绍

研究设计有广义与狭义之分。广义的研究设计在研究方法论的单元中,应告知该研究使用何种方法进行研究,是观察研究还是调查研究,是历史探讨法还是文献探讨法,是评估研究还是再次研究等。任何研究都有必要在研究法这个单元中跟读者介绍所使用的研究法。狭义的研究设计则指,一个研究若使用真实实验方法(the true experimental design)或准实验方法(quasi-experimental design)时,这两种类别有很多不同的设计,也称为研究设计。社会科学中,使用真实实验方法或准实验方法的不多,但是实验设计的原则与重点却可对研究者有莫大的助益。例如,卓春英的研究中,探讨医务社工在医院中的功能时,认为病人在住院时若经过医务社工的处遇,其在医院里的适应情况会比未经过医务社工处遇的为佳。在她的研究方法论中,就简单地画出了这个研究设计图:

$$\begin{array}{ccc} R & \mathrm{X} & O1 \\ R & & O2 \end{array}$$

R 代表经验随机抽样,X 是实验组的病患,$O1$ 则指在进医院就医时曾接受社会工作的处遇后的观察值(即实验的效果),$O2$ 则指第二组控制组并没有经过处遇的观察值(控制组的结果)。严格说来,上述的图示是真实实验中的研究设计之一,卓春英的研究虽非实验研究,但把这个图示画出后,读者马上就可以知道她研究的大概状况。

四、研究假设的说明

研究的目标就是要证明假设是否成立,纵使在问题叙述或文献探讨中或多

或少都谈过研究的概要，也提过研究的假设与方向，但是在研究方法论中，必须把研究的假设“正式”写出来。而交代这些假设时，研究者经常会面临主要假设与次要假设的问题。若研究假设中的主要假设是：

> 父母的婚姻品质会影响子女的婚姻满意度，父母的婚姻品质愈高，其子女的婚姻满意度也必愈高。

乍看之下，这个假设既清楚又简单，但是实际进行时，马上要考虑到什么是父母亲的婚姻品质？什么又是子女的婚姻满意度？若父母亲婚姻品质的操作定义是：(1) 父母亲之间的婚姻满意度；(2) 经济情况满意度；(3) 社交、子女，以及对未来生活的满意度。子女的婚姻满意度若再分为：(1) 生活满意度；(2) 沟通品质满意度；(3) 家庭生活内涵满意度……四五项，显然，一个一个变量交叉分析，一定是错综复杂，若再碰上某些变量有显著相关，某些变量却又无显著相关，那该假设到底是成立还是不成立？此时研究者可能就要大伤脑筋。所以为了让研究的概念更清楚，在陈述假设时，一般都会在主要假设下把次要假设也一个一个写出来，以便在做完分析时马上就可以知道该次要假设是成立或是不成立。而当一些次要假设的成立与否都一目了然时，主要假设到底要如何说明，研究者也会因着这些次要假设的状况心中有数，在说明上就不会产生太大的问题。

五、研究变量的操作性定义

假设既已交代清楚，假设中所牵涉的概念或变量就必须一一说明，把概念或变量到底是什么意思讲得具体、清楚，此即为该概念或变量的操作性定义。

例如，“沟通品质”的操作性定义，可能需要受访者接受“沟通量表”检定后得到实际分数，“父母亲的婚姻品质”可能需要研究者一个一个项目地进行访问与调查，也可能纯粹使用某个现成量表而已。总而言之，把一个研究的概念与变量变成可测量、可观察，甚至可以用某些数字来说明的过程就是操作化。许多研究的变量名称可能大同小异，但是研究者必须叙述出自己的操作性定义与别的研究之间的差异所在，否则难免会造成对事实认知的偏差。在许多教育学或心理学的研究论文当中，操作性定义经常会以某种量表所测量的分数来呈现。

六、量表的制作过程

量表是量化研究中最重要的工具,如何“算”出实验组与控制组之间有差别?必须靠量表来测量。如何证明方案确实有效?而且已经达到预期的效果?也须靠量表来测量。没有量表,量化研究便无伸展的余地。如果量表是那么重要,研究者就必须事先“证明”他的研究所使用量表的可信度与有效度,也因此,整个量表的制作过程就必须先做交代。

如何陈述有关量表制作的内容呢?一般而言,研究者应先阐述量表题目的来龙去脉。研究的理论基础是什么?这些理论如何衍化为一些命题或概念?这些概念又转变为哪些自变量与因变量?每个变量又形成哪些量表的题目?每个量表各有几题?如果把这些由理论而题目的过程都交代清楚时,整个研究在内容效度和建构效度的解释上就不会太吃力。

量表的试测过程也必须做详细的说明。试测之前一共有几个题目?研究者试测时用了几个样本?什么样的样本?试测样本资料收集后,用了哪种方法去计算题目的辨别力(DP 值)与 CR 值?Cronbach α 值又如何?根据这些值,题目是如何取舍的?舍了几题?修正了几题?如何修正的?

若是研究者对因素分析与回归分析的使用非常纯熟,也可以借之算出每个题目或每个单元的题目的信度与结构效度,也可以使用“转轴”的方式,重新组合每个单元的题目,若 8 个题目就可以说明一个概念的 95%,那研究者大可不必用 25 道题来测量受访者。题目的重新组合,有时也可以协助研究者对变量重新命名,有时也可能因为这样使整个研究进入一个研究者本来所未预想的境界。

量表说明的重点是把量表的信度与效度报告清楚。从量表的制作过程到试测的过程与结果,需把所可能牵涉的种种信度与效度一一澄清,并设法使量表“无可指责”,这才是量表制作的最高艺术。

七、抽样

样本的内容与抽样的过程当然也要做说明。此时所要说明的是抽样架构与抽样规划及其样本数的确定。整个研究如何决定样本来源?调查哪些单位?访问了哪些样本?样本数如何确定?如何从不同的单位中收集到这些样本?抽样的过程是简单随机抽样还是分层抽样?抑或是集丛抽样?还是多种抽样方法的混合?

有效样本是多少？无效样本的情况为何……都应该在这个单元中一一说明。

很多质性研究者在解释样本来源时，不免认为：反正是质性研究，简略地讲述样本来源也没关系……这是错误的。质性研究虽没有经过抽样过程，但更会因研究对象与研究目标的不同而必须在选取样本的妥当性及必要性上做详细的解释。说明样本的选取是以资讯的丰富性而访问，还是纯粹的机构内的现成个案而已？是以同样问题的个案为主还是故意找极端偏差的个案？整个个案的组合又如何？更严谨地，在总数不多的个案中，必要时还得分门别类地说明个案的来源。

八、资料收集与资料分析

用何种方法收集资料取决于使用何种研究方法。质性研究与量化研究截然不同，量化研究中也有很多研究方法。不管是何种方法，应该说明是否有访员调查？多少访员？如何训练？是否邮寄问卷？共寄了几份问卷？如何催收？每次共得到多少样本？是否使用电话访问？如何抽样？如何电话询问？收集资料一共花了多少时日……当研究者对资料的收集过程做了清楚的解释时，可以大大减少别人的疑虑。

资料的分析方法或策略也必须交代。大致说明如用了哪些统计方法？用了哪些电脑软件？详细说明则包括哪些变量用了哪些统计方法，把整个研究所要分析的变量及其所需采用的统计方法都做详细的交代，以使读者一目了然。一些结构完整的研究在尚未收集资料时就已经知道到时会有多少表格，也清楚表格陈列的次序又如何。资料收集结束，交给分析公司去分析时，也可使分析人员清清楚楚地按照指示去写程式、用电脑。

第四节 论文第四章——“分析与发现”的写作

一、样本简介

论文最重要的主体是分析与发现。如何呈现所分析的资料，把研究的发现叙述得清楚是很大的学问，因为牵涉到资料的种类、分析的方法、呈现的方式以及是否要做附加的解释等问题。有时各个学科也有不同的习惯，不一而是。有的学

科倾向以分析时的统计方法为呈现的主轴,有的学科则以假设的呈现次序来介绍分析的资料。按习惯,社会工作领域大都以主题的呈现为主,所以以假设是否得到验证为主,以至在次序上也经常以主题或假设为经纬。但是在分析所有资料之前,应该先把样本状况做个交代,此即所谓的样本简介(profile)。

要验证所研究的变量是否与其他变量有关联,总应把样本是什么先介绍。熟识了样本的特质,才能了解为何一些变量之间有其特别的关系。所以,在进行资料的统计分析之前,需应用次数分配表做好样本简介。通常是把样本的基本资料列在一总表格中,如表 14 - 1 所示:

表 14 - 1 样本简介示例

变 量	样本数 260	百分比(%)	变 量	样本数 260	百分比(%)
性别			结婚经验		
男	128	49.2	第一次	181	69.6
女	132	50.8	第二次	20	7.7
教育			分与离	59	22.7
小学	42	16.2	子女数		
初中	58	22.3	0 人	20	7.7
高中	77	29.6	1 人	94	36.2
大学	43	16.5	2 人	105	40.3
研究生	40	15.4	3 人及以上	41	15.8
地区			收入		
北区	87	33.5	3 万元及以下	22	8.5
中区	71	27.3	3.1 万—7 万元	100	38.5
南区	63	24.2	7.1 万—10 万元	103	39.6
东区	39	15.0	10.1 万—15 万元	16	6.2
年纪			15 万元以上	19	7.2
20 岁及以下	24	9.2			
21—30 岁	47	18.1			
31—40 岁	71	27.3			
41—50 岁	52	20.0			
51—60 岁	40	15.4			
61 岁以上	26	10.0			

从以上的样本简介中，读者可以了解该研究的样本概要，有关样本的性别比例、受教育程度、住宅地区、收入、子女数及年龄等资料，都在这一表格中一目了然。然而，在列出表格以后，应该有一些文字的叙述，在文字叙述中，重点是“告知”样本的大要，并非“分析”样本，大可不必做太多引申性的解释。有些研究者难免会在样本简介中大力着墨于次数分配表中的“发现”，其实，直截了当地介绍样本即可。当然在表格中，一切资料一定要精准、细密，表格的呈现愈美观愈好，表格千万不要分页陈列，最好在一页中就能“刊载”完毕，否则，在对照上会产生困扰。文字说明部分不在一页之内时，读者阅读上或可接受，但表格分页呈现的情况则应尽力避免。另外一个经常被疏忽的问题是，表格与说明之间不要“距离”太远。若是样本简介的单元是在第 34 页，那表格就不应放在论文后面（如 325 页）集中刊出，以免读者一边看论文，一边还得“找”表格；一手“拿”笔，一手还得“按”住资料，读起来不免劳神。

质性研究的样本简介方式与量化研究不同。由于质性研究的样本不多，有关个别差异所引申的内涵也是重要的研究内容，所以质性研究的样本介绍应该郑重其事。它的样本介绍不是介绍样本整体的性别、受教育程度、年龄、地区等属性资料的分布情形，其重点应该放在介绍研究的个体状况上。样本的基本背景除了属性资料以外，每个样本的生活状况、资历、成长特殊经验及在量化研究的“干扰”因素，在质性研究中都可能成为宝贝。如果在资料分析之前，每个样本能有两三百字的简介，对资料分析的说服力必可提高不少。至于这些个案的简介是否一定要放在正文中，这点上研究者有不同的意见。有人认为放在研究论文之后的附录未尝不可；有的质性研究则宁愿把样本简介放在资料分析的开始部分，认为这样读起来较顺畅，但缺点是会因篇幅太多造成编排上的不均衡……孰优孰劣，视现实情况而定。

二、分析与发现（findings）

资料经过统计分析后，有些资料呈现显著差异，符合研究假设的期待，这就是所谓研究的发现。把这些发现详加叙述是研究最有成就的时刻，但是如何呈现，也应小心翼翼。

从次序上来看，把发现依假设的陈述次序呈现，比依统计分析方法的次序呈现要合理，在阅读上也会较顺畅。毕竟我们不需要告知读者统计的方法如何

应用,而是陈述研究的发现。

统计分析时难免会使用很多表格。表格的陈列应以标准方式为之,社会科学的研究报告中,尤其是APA的格式里,表格均以横线为主,鲜少有直线者,想必是因为考虑到电脑的使用方便所致。表格尽量不要跨页,每个表格陈列后都应有文字的解释,说明表格在统计上的显著意义,使读者了解表格的内容。不同的统计方法,表格的陈列亦有不同,兹把一般最常见的统计表格及统计图简介如下,以供读者们参考(见表14-2、表14-3、表14-4、表14-5和图14-2、图14-3)。

表14-2　平均数、标准差、样本数在两组比较时的呈现方式(年纪不同在受测时感受难度程度的差异)

难　度	年轻组			老年组		
	M	*SD*	*n*	*M*	*SD*	*n*
低难度	0.05	0.08	12	0.14	0.15	18
中难度	0.05	0.07	15	0.17	0.15	12
高难度	0.11	0.10	16	0.26	0.21	14

M 为平均数;*SD* 为标准差;*n* 为样本数。

表14-3　相关系数的呈现方式(不同年龄组的生涯规划程度与适应层次间的相关分析)

Subscale	1	2	3	4
学生组(*n*=200)				
1. Tranquillity	—	0.93	−0.09	0.73
2. Goodwill		—	−0.34	0.62
3. Happiness			—	0.14
4. Elation				—
成人组(*n*=189)				
1. Tranquillity	—	0.42	−0.07	0.52
2. Goodwill		—	−0.43	0.62
3. Happiness			—	0.47
4. Elation				—

1,2,3,4为不同的生涯规划程度,1最低,4最高;适应层次则分为平静、安好、幸福等;不同年龄层则只分为学生组与成人组两组。

表 14－4 单因子变异数分析表的呈现方式(古典制约的变异数分析)

Source	df	F	
		Finger *CR*	Irrelevant *CR*
Between subjects(组间)			
焦虑(*A*)	2	0.76	0.26
震惊(*S*)	1	0.01	0.81
A * *S*	2	0.18	0.50
S within-group error	30	(16.48)	(15.73)
Within subjects(组内)			
Blocks(*B*)	4	3.27**	4.66**
B * *A*	8	0.93	0.45
B * *S*	4	2.64*	3.50**
B * *A* * *S*	8	0.58	0.21
B * *S* within-group error	120	(1.31)	(2.46)

说明：( )指平均数平方误差；*CR* 指制约反应；*B* 指因变量。
* $p<0.05$；** $p<0.01$

表 14－5 影响女儿对奉养父母信念的回归分析表(用逐步回归的方式)

Variable	*B*	*SE B*	β
Step 1			
女儿受教育程度	−5.89	1.93	−0.41*
母亲年龄	0.67	0.31	0.21*
Step 2			
女儿受教育程度	−3.19	1.81	−0.22
母亲年龄	0.31	0.28	0.14
对年老者的态度	1.06	0.28	0.54*
关怀与感受	1.53	0.60	0.31*
Dogmatism	−0.03	0.10	−0.04

$R^2=0.26$ for Step 1；$R^2=0.25$ for Step 2
* $p<0.05$

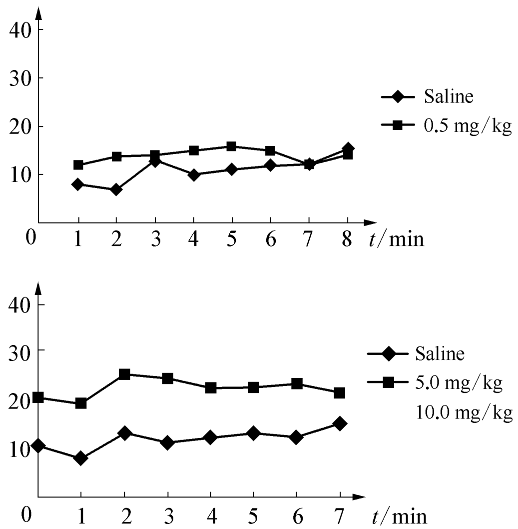


图 14-2 线条图形

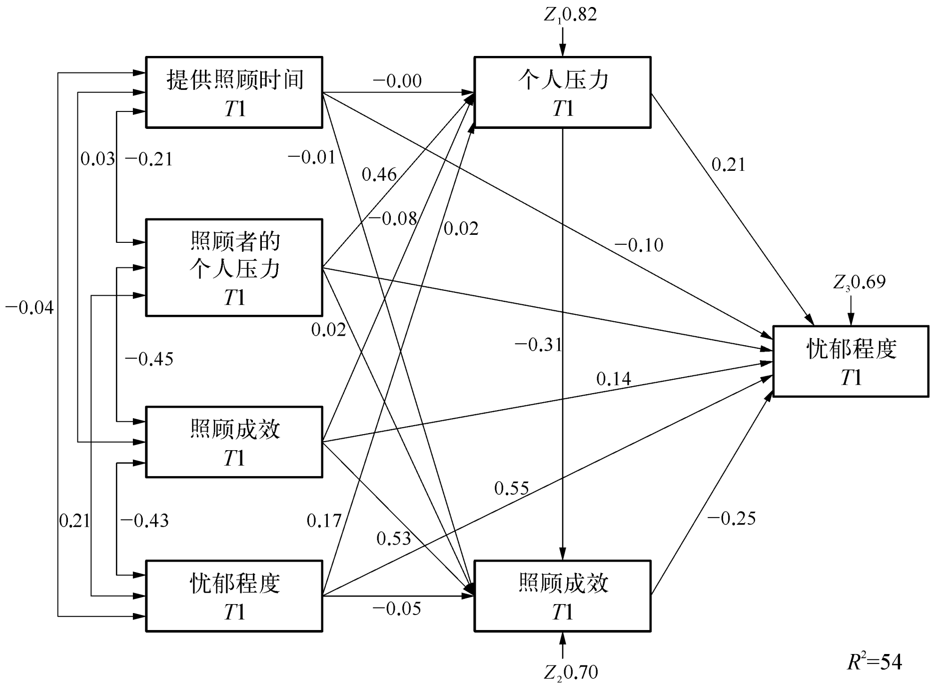


图 14-3 路径模式的图形呈现(path model)

说明：路径分析模式侧重于分析照顾时间的长短、主要照顾者的压力，以及主要照顾者的照顾成效对受照顾者的忧郁所产生的影响程度。

在不少调查研究中，常借着电脑计算得出表格。这些表格有的在统计上显

著，有的不显著。在统计上显著的表格应该呈现，并且加以解释；而如何处理不显著的表格却是学问，总的来说，多以研究的性质或是表格的多寡而定。当然，不显著的表格用不着费心解释变量之间的关系，但是若有系统上的不显著，而且与主题有微妙关联的话，若能以综合的表格稍加说明，资料的呈现会更清楚。

在统计上呈现显著的表格下方，应有文字上的说明，但文字说明应该详细到什么程度也有不同的说法，有的研究者认为只要说明表格的“内容”即可，不要做太多的引申解释，而应把所有的较为详细的解释全部放在另外的单元中（如“讨论”的单元）。有的研究者则认为，表格下方用文字解释了统计上的意义以后，若能再加以意义上的说明则更是一气呵成，能使读者对表格有更深一层的了解，在阅读时也会较为自然顺畅。

第五节　论文第五章——“结论”的写作

事实上在每个章节后面都应该有“结论”。小章节有小结论，大章节有大结论。随着结论的层次不同，抽象的层次也有不同，结论绝对不是把前面所言再重复说明一次而已。结论告一段落时，研究者不妨把论点拉回研究方法论里所谈及的假设，把结论引申到研究的假设是否已得到证明（或排斥），如此一来，整个研究便可以层次分明，而愈到后面，整个研究的结果如何便可呼之欲出了。

各章节虽然都有结论，但在论文的第五章仍然是以“结论”为主轴。在这一章中，大体可包括结论（conclusion）、应用与建议（implication & suggestion）、研究限制（limitation）等部分，兹分别说明如下。

一、结论

此时的结论是总结性的结论，有别于各章节的小结论。当然有些研究者会趁机把各章节的结论再稍微浓缩一下，在此单元中再做一次介绍。其实，在结论这章，其内容的抽象层次应较前面章节的内容为高，前面的章节以某个变量或某一假设的相关资料作为结论的焦点，但是结论若在本章出现的话，则应是综合性的结论，所考虑的应该是通盘的变量与整体性（set）的假设，反正愈到后

面,抽象层次必然愈高,所考虑的主题更是愈广。

二、应用与建议

从社会工作的角度来看,社会工作研究有别于其他学科的研究,最大的不同是“研究应用”的极度强调。社会工作既然是个专业,它与日常生活便息息相关,“实用性”最高。因此,研究是从专业中来,研究也应该回到专业中去,研究的种种成果与发现,理所当然必须对专业服务水准有提升作用,否则就不是社会工作研究。

在建议的单元中,研究者常遇到的问题是把建议当“作文”来写。仿佛抓到千载难逢的发泄机会一样,将一些平日的想法拼命做文章。其实科学的精神是“有几分证据才说几分话”,研究要有多少“建议”取决于研究有多少“发现”。因此在建议时,每提一个建议前不妨都先说出是根据研究中的哪一个发现而来,再从这些研究发现中逐一“做文章”,如此一来,整个研究的严谨性必可提高。

社会工作专业的特性是有很强烈的实务本质。社工的硕博士论文若只在学理上钻研,而疏忽应该对社工实务有所帮助的话,就会与社会学或心理学的论文一样,不能展现社工研究的特色。所以社工研究在“建议”的写作上应该相对强化,而且必须与实务挂上钩。建议社工研究者应该设法把研究的每一个发现都引申到具体的实务上。所以建议的每一点或每一段应该是以“根据研究的 xxx 发现,建议在 yyy 的实务场合中应该做 zzz 的配合或改变”为基本的模式,当社工研究能如此强调在实务中的应用时,相信对国内的社工实务会有很大的助益。

三、研究限制

把整个研究中所没有办法克服的问题与缺点一一交代,是研究者的义务。因整个研究中的弱点只有研究者本人最为清楚,为免使后来的研究者重蹈覆辙,在研究限制的单元中,应该好好阐述清楚。除了一般研究过程常见的困扰与缺点以外,研究的哲学困境、研究应用时的难题与缺陷也不妨一并说明,也可根据这些限制与说明,对未来的研究者做建议。

第六节　参考文献与附录

把研究做到资料分析，可以说是“大功告成”，写完了结论、建议与研究限制以后，更可以说是“功德圆满”。一般研究者到这个阶段，也经常以为可以收拾资料，准备上餐厅好好庆祝一下，无心再做太多的修补与增订。其实，论文的最后阶段是最重要的时刻，尤其是有关参考文献与附录的交代，万万不可忽略。有经验的读者在看一篇论文之前，习惯性地一定会先翻阅该论文的参考文献，借着参考文献可以知道该论文究竟引用了哪些资料，有没有承接该主题的重要作者（或大师）的内容，该资料是否够新（合乎潮流），该资料有否偏向于某些方向而忽略了其他重要的单元……在未读论文内容之前，若先把参考文献过目，大体上就可以知道论文的内涵。若以严格的眼光来看，从参考文献的格式是否合乎“规格”更可知道研究者处理资料的态度。社会工作者平日忙于服务别人，认真做好实务的工作；从事研究时，虽然也在前面的阶段付出努力，下了相当大的功夫，然而却在参考文献部分潦草作业，因而坏了整篇论文的严谨度，相当可惜。

有关文后参考文献的格式，社会科学界的研究者们多以美国心理师学会（APA）所颁布的格式为主。各种相关书籍都对 APA 参考文献与注解的格式有详细介绍，读者可以参考。

中文参考文献与英文参考文献在中文的论文中经常都是分开陈列。许多研究者喜欢在各个参考文献之前还加上阿拉伯数字来区分，事实上大可不必，因为英文的资料已经按字母排列，中文的参考文献也多以姓氏笔画的多寡来排列，所以大可不必加上数字号码来区分（参考文献前加上数字的话，反而把 APA 的原有格式弄乱了）。参考文献分中英文排列业已足够，有的研究者还喜欢再分成专著类与期刊类等，以为分得愈精细愈好，事实却不尽然。因为读者在阅读论文时，从注解所得到的资讯只有作者的姓、年代与页数，并不知道是期刊或是专著还是杂志，翻到参考文献要找寻出处时，却还要花很多的精力在不同的类别中寻找，可见该分类只是画蛇添足，徒增困扰。

在英文参考文献中，很多研究者无法判定原作者的姓（last name or

surname)与名(first name)的问题。参考文献中作者是以姓来呈现,名字则缩写即可。然而一般在读论文时,作者的姓名经常以名在前姓在后的方式呈现,作者在写参考文献时不可不慎。在一般的状况下,作者的名字呈现是以姓开始时,应马上会加上逗点“,”,所以若看到姓名中有逗点时,前面那个单字应该是作者的姓——这是简单的判断方式。

更需注意的是,在论文中引用了哪些资料,在参考文献(参考资料)的单元中也只能列出这些资料。有些人以为参考文献列得愈多愈好,所以不管三七二十一,拼命把所有的相关专著或期刊资料都列进去,好让别人误以为你真的看了所有的文献,这并不是诚实的研究者应该有的做法。有经验的论文评论者在看研究论文时,当他看到一个被引用的作者,会抽样式地马上翻到参考资料部分,去查查这个人有没有被列入参考文献(用这个方法可以探知写作的人的研究态度是否严谨);另外一个方式是看到后面参考文献的作者名字,然后回头去看你在论文的哪个部分用了这个作者的文献。若论文的写作能通过这些考验,评论者对研究者则会有较佳的“印象”,当然其他分析的细节或角度可能就会比较通融也说不定。

附录(appendix)的用意是把不能放在正文中却又必须向读者交代的资料放在论文的后面,供读者参考。一般是该论文的问卷、进行研究时的重要书信、调查时的访问人员名单,还有有关的量表等,质性研究的样本简介也经常放在附录里。

第七节　其　　他

论文的呈现还需考虑到封面、摘要、致谢词、论文目录、表格目录,以及图表目录等。

关于封面的格式,教育主管部门已有定规,举凡学校名称、论文名称、指导教授、学生姓名及日期等,均需依其规定书写,在该页的哪处位置也应考究。毕竟大家都遵守其规范,每个人的资料整理与置放等才不致零乱。

摘要依外文的规定都不能超过300字,必须在单页内完成,原因是为了制作时的格式能够统一。中文论文的很多摘要都是七八百字,横跨两页,装订时

固然不会有所困扰，但是未来在整理为电子文档，或是要出版论文时，一定会碰到困难，研究者不妨从现在就开始注意。

致谢词固然可以让研究者各抒所感，但是篇幅仍应考虑，避免拉杂，否则违背了致谢的真意。论文一旦完成，要感谢的人当然很多，但若都纳入致谢词也不免烦琐，而且致谢的性质反而分散不显著。相反地，有时只列出真正重要非致谢不可的，或是简单的感言，反而较为亲切实在。

论文印刷若是讲究一点，很容易讨好读者。视窗及 Word 的软件应用，使一般的研究者有能力依自己的能力安排论文印刷时的“布局”与“格调”，从字形的选择、表格的编排、空间的使用，以及图形的创作……若在印刷时亦能细加规划，则论文出炉时一定会让人爱不释手。其实印刷时最重要的不仅是外表的美观而已，而是是否经过严格的校正，不让任何因为疏忽所造成的错误出现在论文中。这是研究者在论文出版时应有的自我要求。与读者们共勉。

习　题

1. 研究报告的撰写在结构上应注意哪些方面？
2. 请选择一个你有兴趣的社会问题，并用社会工作研究法的角度，以量化的方法来探讨。
   (1) 请写出与此问题有关的研究问题。
   (2) 请依此研究题目写出主要假设。
   (3) 请指出此假设的自变量、因变量及中介变量。
   (4) 请写出你的研究设计，并用例子指出该研究设计可能面临的缺点。
   (5) 请写出你的研究计划，并指出它的可行性。

参考文献

一、中文部分

王保进.1999.视窗版 SPSS 与行为科学研究.台北：心理出版社.

王丛桂，罗国英.1992.社会研究的资料的处理.台北：黎明文化事业公司.

白秀雄.1986.社会工作.台北：三民书局股份有限公司.

朱经明.1989.教育统计学.台北：五南图书出版公司.

李沛良.1987.社会研究的统计分析.武汉：湖北人民出版社.

欧沧和，李茂能.1985.社会科学研究法辞典.高雄：复文图书出版社.

林清山.1992.心理与教育统计学.台北：东华书局股份有限公司.

邱皓政.2009.量化研究与统计分析：SPSS 中文视窗版资料分析范例解析.重庆：重庆大学出版社.

陈向明.2002.社会科学质的研究.台北：五南图书出版公司.

杨志良.1983.生物统计学新论.台北：巨流图书公司.

颜月珠.1989.统计学.台北：三民书局股份有限公司.

罗积玉.1990.多元统计分析：方法与应用.台北：科技图书公司.

Robert L. Miller，et al. 2006. SPSS 在社会科学的应用.庄文忠，译.台北：五南图书出版股份有限公司.

二、英文部分[①]

Allen-Meares, P. & Lane, B. A. (1987). Grounding Social Work Practice in Theory: Ecosystems. *Social Casework*, 68, 515-522.

① 为方便读者查阅，本书按原版复制英文参考文献。

American Psychological Association. (1983). *Publication Manual of the American Psychological Association* (3rd edition). Washington, D.C.

American Psychological Association. (1987). *Casebook on Ethical Principles of psychologists.* Washington, D.C.

Asher, H.B. (1976). *Causal Modeling*. Beverly Hills, Calif.: Sage.

Assael, H., & Keon, J. (1982). Non-sampling vs. Sampling Errors in Survey Research. *Journal of Marketing,* 46, 114-123.

Babbie, E. (1986). *Observing Ourselves: Essays in Social Research.* Belmont, A: Wadsworth.

Babbie, E. (1990). *Survey Research Method.* Belmont, CA: Wadsworth.

Babbie, E. & Rubin, A. (1992). *Research Methods for Social Work* (2th ed.). Pacific Grove, California.

Babbie, E. R. (1973). *Survey Research Methods*. Belmont, Calif.: Wadsworth.

Baird, D.C. (1988). *Experimentation: An Introduction to Measurement Theory and Experiment Design* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Baker, T.L. (1994). *Doing Social Research.* McGraw-Hill.

Baltes, P. B., Cornelius, S. W., & Nesselroade, J. R. (1979). Cohort Effects in Developmental Psychology. In J. R. Nesselroade & P.B. Baltes(Eds.), *Longitudinal Research in the Study of Behavior and Development.* New York: Academic Press.

Bandler, R. & John G. (1975). *Patterns of the Hypnotic Techniques*. Milton H. Erikson, M. D.,Vol. 1 Cupertino, CA: Meta.

Baumrind, D. (1985). Research Using Intentional Deception: Ethical Issues Revisited. *American Psychologist*, 40, 165-174.

Beach, D. P. & Alvager, K. E. (1992). *Handbook for Science and Technical Research*. Engelwood Cliffs, NJ: Prentice-Hall.

Beck. E. M. & Tolnay, S. E. (1990). The Killing Field of the Deep South: The Marker for Cotton and the Lynching of Blacks. 1882-1930. *American Sociological Review*, 55, 526-539.

Becker, H. S. (1958). Problems of Inference and Proof in Participant

Observation. *American Sociological Review*, 23, 652-660.

Berelson, B. (1954). "Content Analysis" in Gardner Lindzey (ed.), *Handbook of Social Psychology,* Wesley, Cambridge, Mass.

Bergin, A. E. & Lambbert, M. J. (1978). The Evaluation of Therapeutic Outcomes. In S. L. Garfield and A. E. Bergin (Eds.), *Handbook of Psycho therapy and Behavior Change: An Empirical Analysi*s (pp. 239-252). New York: John Wiley.

Berry, W. D. (1984). *Non-recurvisive Causal Models.* Beverly Hills, Calif.: Sage.

Berry, W. D., & Feldman, S. (1985). *Multiple Regression In Practice*. Beverly Hills, Calif.: Sage.

Bertera, E. M., & Bertera, R. L. (1981). The Cost-Effectiveness of Telephone vs. Clinic Counseling for Hypertensive Patients: A Pilot Study. *American Journal of Public Health*, 71, 626-629.

Bettleheim, B. (March 1, 1982). Reflections: Freud and Soul. *New Yorker.* 52.

Beveridge, W. I. B. (1950). *The Art of Scientific Investigation*. New York: Vintage Books.

Beveridge, W. I. B. (1957). *The Art of Scientific Investigation Revised*. New York: W. W. Norton.

Beveridge, W. I. B. (1980). *Seeds of Discovery*. New York: W. W. Norton.

Blalock H. M. (1972). *Social Statistics.* McGraw-Hill Book Company.

Bloombaum, M. (1983). The Hawthorne Experiments: A Critique and Reanalysis of the First Statistical Interpretation by Franke and Kaul. *Sociological Perspective*s, 26, 71-88.

Bonjean, C. M., Hill, R. J., & Mclemore, S. D. (1967). *Sociological Measurement: An Inventory of Scales And Indices*. San Francisco: Chandler.

Borgatta, E. F., & Bohrnstedt, G. W. (1980). Level of Measurement: Once Over Again. *Sociological Methods and Research,* 9, 147-160.

Borman, K. M. & Judith, P. G. (1986). Ethnographic and Qualitative Research Design and Why It Doesn't Work. *American Behavioral*

Scientist, 30(1), 42-57.

Box, G. E. P., & Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Ray.

Bradburn, N. M. & Sudman, S. (1988). *Polls and Surveys: Understanding What They Tell Us*. San Francisco: Jossey-Bass.

Brandt, R. (1972).*Studying Behavior in Natural Settings*. New York: Holt, Rinehart & Winston.

Brinkerhoff, R.O. & Dressler, D.E. (1990). *Productivity Measurement.* (生产力之衡量.王昭正译于2001).台北: 弘智.

Browne, A. (1987). *When Battered Women Kill.* New York: Free Press.

Brownlee, K. A. (1975). A Note on the Effects of Non-response on surveys. *Journal of the American Statistical Association*, 52(227), 29-32.

Bruyn, S. (1963). The Methodology of Participant Observation. *Human Organization*, 21, 224-35.

Bruyn, S. (1966). *The Human Perspective in Sociology: The Methodology of Participant Observation.* Englewood Cliffs, NJ: Prentice-Hall.

Campbell, D. & Stanley, J. (1963). *Experimental and Quasi-Experimental Designs for Research*. Chicago: Rand McNally.

Campbell, D. T. (1975). Reforms as Experiments. In E. L. Struening and M. Guttentag (Eds.), *Handbook of Evaluation Research.* Vol. 1. Beverly Hills, Calif.: Sage.

Campbell, D. T., & Erlebacher, A. (1970). How Regression Artifacts in Quasi-experimental Evaluations Can Mistakenly Make Compensatory Education Look Harmful. In J. Hellmuth (Ed.), *Disadvantaged Child*, Vol. 3: Compensatory Education: A National Debate. New York: Bruner/Mazel.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and Quasi-experimental Designs for Research Skokie,* I11.: Rand McNally.

Ceci, S. J., Peters, D., & Plotkin, J. (1985). Human Subjects Review, Personal Values, and the Regulation of Social Science Research. *American Psychologist,*40, 994-1002.

Chavetz, J. (1978). A *Primer on the Construction and Testing of Theories in Sociology*. Itasca, IL: Peacock.

Cohen, L., & Manion, L. (1985). *Research Methods in Education*. London: Croom Helm.

Committee for the Protection of Human Participants in Research. (1982). *Ethical Principles in the Conduct of Research with Human Participants*. Washington, D.C.: American Psychological Association.

Conner, R. F. (1977). Selecting a Control Group: Analysis of the Randomization Process in Twelve Social Reform Programs, *Evaluation Quarterly*, 1, 195-244.

Cook, T. D. & Campbell, D. T. (1979). *Quasi-experimentation: Design and Analysis Issues for Field Settings. Skokie*, I11.: Rand McNally.

Cook, T. D., Dintzer, L., & Mark, M. M. (1980). The Causal Analysis of Concomitant Time-series. *Applied Social Psychology Annual*, 1, 93-135.

Cooper, H. M. (1989). *Integrating Research*. (*研究文献之回顾与整合*.高美英译于1999).台北：弘智.

Cozby, P. C. (1984). *Using Computers in the Behavioral Sciences*. Palo Alto, Calif.: Mayfield.

Crabtree B. & Miller, W. L. (1999). *Doing Qualitative Research*. (*质性方法与研究*.黄惠雯、童琬芬、梁文蓁、林兆卫等译于2002).台北：韦伯.

Craig, P. (1978). *The Heart of a Teacher: A Heuristic Study of the Inner World of Teaching*. Ann Arbor, MI: University Microfilms International.

Cronbach, L. J., & Meehl, P. E. (1955). Construct Validity in Psychological Tests. *Psychological Bulletin*, 52, 281-302.

Curry, T. J. & Alfred, C. C. (1977) *Introducing Visual Sociology*. Dubuque, Iowa: Kendall/Hunt.

Dabbs, J. M., Jr. (1982). Making Things Visible. In J. Van Maanen, J. M. Dabbs, Jr., & R. R. Faulkner(Eds.), *Varieties of Qualitative Research* (pp.97-123).Beverly Hills, CA: Sage.

Das, M. & Giri, N. (1986). *Design and Analysis of Experiments*. 2nd ed.,

New York: John Wiley & Sons.

David P. B. & Torsten K. E. A. (1992). *Handbook for Scientific and Technical Research.* Englewood Cliffs, NJ: Prentice-Hall.

DeMaria, W. (1981). Empiricism: An Improvised Philosophy for Social Work Research. *Australian Social Work,* 34, 3-8.

Delbecq, A.L., Van de Ven, A.H., & Gustagson, D.H. (1975). *Group techniques for program planning: A guide to nominal group and Delphi process.* Glenview, IL: Scott, Foresman and Co.

Denzin, N. K. (1978). *The Research Act: A Theoretical Introduction to Sociological Methods.* New York: McGraw-Hill.

Denzin, N. K. (1978). *The Logic of Naturalistic Inquiry in Sociological Methods.* New York: McGraw-Hill.

Denzin, N.K. (1989). *Interpretive Interactionism.* (*解释性互动论.*张君玫译于1999).台北：弘智.

De Vaus, D. (2001). *Research Design in Social Research.* London: Sage.

De Vellis, R.F. (1991). *Scale Development.* (*量表的发展：理论与应用.*吴齐殷译于1999).台北：弘智.

Dillman, D. A. (1978).*Mail and Telephone Surveys: The Total Design Method.* New York: John Wiley & Sons.

Donald, M. N. (1960). Implications of Non-response for the Interpretation of Mail, *Quarterly*, 24(1), 99-114.

Dubin, R. (1978). *Theory Building.* New York: Free Press.

Dunkelberg, W. C., & Day, G. S. (1973). Non-response Bias and Callbacks in Sample Surveys. *Journal of Marketing Research,* 10, 160-168.

D'Youville College. *Measures: Non-reactive Research in Social Science.* Chicago: Rand McNally.

Emerson, R. M. (ed.). (1983). *Contemporary Field Research.* Boston: Little Brown.

Epstein, I. (1985). Quantitative and Qualitative Methods. In R. Grinnell, Jr.(Ed), *Social Work Research* (pp.263-274). Itasca, IL: F. E. Peacock

Publishers.

Feick, L. F. (1989). Latent Class Analysis of Survey Questions that Inc' Don't Know Responses. *Public Opinion Quarterly*, 53, 525-47.

Fetterman, D. M. (1988). Approaches to Evaluating Education. *Educational Researcher*, 17(8), 17-23.

Fetterman, D. M. (1988). *Qualitative Approaches to Evaluation in Education: The Silent Scientific Revolution*. New York: Praeger.

Fetterman, D. M. (1989). *Ethnography: Step by Step*. Newbury Park, CA: Sage.

Fienberg, S. (1977). *The Analysis of Cross-classified Categorical Data*. Cambridge, Mass.: MIT Press.

Finsterbusch, K., & Motz, A. B. (1980). *Social Research for Policy Decisions*. Belmont, Calif.: Wadsworth.

Fiske, D. W. (1987). Construct Invalidity Comes From Method Effects. *Educational and Psychological Measurement*, 47, 285-307.

Franke, R. H., & Kaul, J. D. (1978). The Hawthorne Experiments: First Statistical Interpretation. *American Sociological Review*, 43, 623-643.

Frankfort-Nachmias C. & Nachmias, D. (1996). *Research Methods in the Social Sciences*. (社会科学研究方法.潘明宏、陈志玮合译).台北：韦伯.

Freeman, D. (1983). *Margaret Mead and Samoa: The Making and Unmaking of an Anthropological Myth*. Cambridge, MASS.: Harvard University Press.

Geismar L. L. & Wood, K. M. (1982). Evaluating Practice: Science as Faith. *Social Casework*, 63, 266-272.

Glock, c. (ed.) *Survey Research in the Social Science* New York: Russell Sage Foundation.

Golden, M. P. (ed.). (1976). *The Research Experience*. Itasca, IL: Peacock

Goldman R. N., & Weinberg, J. S. (1985). *Statistics: An Introduction*. Englewood Cliffs, N. J.: Prentice-Hall.

Goldstein, H. (1981). *Social Learning and Change: A Cognitive Approach to*

Human Service. S. C., Columbia: University of South Carolina Press.

Goldstein, M. & Goldstein, I. F. (1979). *How We Know*. New York: Plenum Press.

Gooding, D., Pinch, T. & Schaffer, S. (eds.). (1989). *The Use of Experiment*. Cambridge University Press.

Grady K.E. & Wallston, B.S. (1988). *Research in Health Care Settings*. (*医疗保健研究法*.赖文福译于2000).台北：弘智.

Green, B. F., & Hall, J. A. (1984). Quantitative Methods for Literature Reviews. *Annual Review of Psychology*, 35, 37-53.

Guba, E. G.. & Yvonna, S. L. (1981). *Effective Evaluation: Improving the Usefulness of Evaluation Results Through Responsive and Naturalistic Approaches*. San Francisco: Jossey-Bass.

Guba, E. G. & Yvonna, S. L. (1988). Do Inquiry Paradigms Imply Inquiry Methodologies? Pp.89-115 in *Qualitative Approaches to Evaluation in Education: The Silent Scientific Revolution,* edited by D. Fetterman. New York: Praeger.

Guba, E. G. (1978). *Toward a Methodology of Naturalistic Inquiry in Education Evaluation*. CSE Monograph Series in Evaluation no. 8. Los Angeles: University of California, Los Angeles, Center for the Study of Evaluation.

Guenzel, P. J., Berckmans, T. R., & Cannell, C. F. (1983). *General Interviewing Techniques: A Self-instructional Workbook for Telephone and Personal Interviewer Training*. Ann Arbor, Mich.: Institute for Social Research.

Hakim, C. (1982). *Secondary Analysis in Social Research*. Allen and Unwin, London.

Hawkins. C. & Sorgi, M. (eds.). (1985). *Research*. New York: Springer-Verlag.

Haworth, G. O. (1984). Social Work Research, Practice, and Paradigms. *Social Service Review*, 58, 343-357.

Hedrick, T.E., Bichman, L. & Rog, D.J. (1993). *Applied Research Design-A*

Practical Guide. New Park, California: Sage.

Heineman, M. B. (1981). The Obsolete Scientific Imperative in Social Work Research. *Social Service Review,* 55, 371-397.

Heise, D. R. (1975). *Causal Analysis*. New York: John Wiley.

Hirschi, T. & Selvin, H. (1973). *Principles of Survey Analysis*. New York: Free Press.

Hirsh, S. K. & Jean, M. K. (1987). *Introduction to Type in Organizational Settings*. Palo Alto, CA: Consulting Psychologists.

Holsti, O. R. (1969). *Content Analysis for the Social and Humanities Reading*. Mass.: Addison-Wesley.

Homans, G. C. (1971). Reply to Blain. *Sociological Inquiry*, 41, 23.

Hoover, K. & Donovan, T. (1995). *The Elements of Social Scientific Thinking*.(社会科学方法论的思维.张家麟译于2000).台北：韦伯.

Hunt, M. (1985). *Profiles of Social Research: The Scientific Study of Human Interactions*. New York: Basic Books.

Hyman, H. (1972). *Secondary Analysis of Sample Surveys*. New York: John Wiley & Sons.

Jones, S. R. G. (1990). Worker Independence and Output: The Hawthone Studies Revaluated. *American Sociological Review,* 55, 176-190.

Jorgensen, D.L. (1989). *Participant Observation*. (参与观察法.王昭正与朱瑞渊合译于1999).台北：弘智.

Kazi, M.A.F. & Wilson, J. (1996) Applying Single-Case Evaluation in Social Work. *British Journal of Social Work*, 26, 699-717.

Kiecolt, E. J. & Nathan, L. E. (1985). *Secondary Analysis of Survey Data.* Beverly Hills, CA: Sage Publications.

Kim, J., & Mueller, L. W. (1978). *Introduction to Factor Analysis*. Beverly Hills, Calif.: Sage.

Kimmel, A. (1988). *Ethnics and Values in Applied Social Research*. New Bury Park, CA: Sage.

Kimmel, A.J. (1988). *The Ethics and Values in Applied Social Research*. (应

用性社会研究的伦理与价值.章英华译于1999).台北：弘智.

Kish, L. (1965). *Survey Sampling*. New York: John Wiley & Sons.

Kraemer, H. C., & Thiemann, S. (1987). *How Many Subjects? Statistical Power Analysis in Research*. Beverly Hills, Calif.: Sage.

Krenz, C. & Gilbert S. (1986). What Quantitative Research Is and Why It Doesn't Work. *American Behavioral Scientist*, 30(1), 58-69.

Kroeger, O. & Janet M. T. (1988). *Type Talk*. New York: Delacorte.

Kuhn, T. S. (1970). *The Structure of Scientific Revolutions*. Chicago.: University of Chicago Press.

Kumar R. *Research Methodology-A step-by-step guide for beginners*.(*研究方法——步骤化学习指南*.潘中道、黄玮莹、胡龙腾合译).台北：学富.

Layder D. (1998). *Sociological Practice: Linking Theory and Social Research*.(*社会科学研究的理论与方法*.简守邦译）.台北：韦伯.

Lazarsfeld, P. & Rosenberg, M. (eds.) (1955). *The Language of Social Research*. New York: Free Press of Glencoe.

Lazarsfeld, P. (1959). Problem in Methodology. In Robert K. Merton(ed.). *Sociology Today*. New York: Basic Books.

Leedy, P. D. (1989). *Practical Research: Planning and Design*. New York: Macmillan.

Lennox, C.E. (1987). *Guidelines for high school suicide prevention program: A Delphi study*. UML Dissertation Services, East Texas State University.

Leviton, L. C., & Hughes, E. F. X. (1981). Research on the Utilization of Evaluations: A Review and Synthesis. *Evaluation Review*, 5, 525-548.

Lincoln, Y. S. & Egon, G. G. (1985). *Naturalist Inquiry*. Newbury Park, CA: Sage.

Linnan. (1976). *Foundations of Social Research*. McGraw-Hill Book Company.

Locander, W., Sudman, S., & Bradburn, N. (1971). *Analyzing Social Settings: A Guide to Qualitative Observation and Analysis*. Belmont, Calf: Wadsworth.

Locander, W., Sudman, S., & Bradburn, N. (1976). An Investigation of Interview Method, Threat and Response Distribution. *Journal of the American Statistical Association*, 71, 269-275.

Lofland, J. & Lofland, L. H. (1984). *Analyzing Social Settings*. Belmont, CA: Wadsworth.

Lofland, J. (1971). *Analyzing Social Setting: A Guide to Qualitative Observation and Analysis*. Belmont, CA: Wadsworth.

Lofland, J. (1984). *Analyzing Social Setting*. Belmont, CA: Wadsworth.

Lowrance, W. W. (1982). *Modern Science and Human Values*. New York: Oxford University Press.

MacRae, D. J. (1976). *The Social Function of Science*. New Haven, CT: Yale University Press.

Magidson, J. (1977). Toward a Causal Model Approach for Adjusting for Preexisting Differences in the Nonequivalent Control Group Situation: A General Alternative to ANCOVA. *Evaluation Quarterly*, 1, 399-420.

Maluccio, A. N. (1979). *Learning From Clients*. New York: Free Press.

Mangione, T.W. (1995). *Mail Survey*. (*邮寄问卷调查*.王昭正、朱瑞渊译于1999).台北：弘智.

Martin, D. W. (1985). *Doing Psychology Experiments*. Monterey, CA: Brooks/Cole.

Mathews, R., Matthews, J. K. & Speltz, K. (1989). *Female Sexual Offenders an Exploratory Study*. Orwell, VT: Safer Society Press.

Maxwell J.A. (1994). *Qualitative Research Design: An Interactive Approach*. (*质化研究设计——一种互动取向的方法*.高熏芳、林盈助、王向葵等译于2001).台北：心理.

Maxwell, S. E. & Delaney, H. D. (1985). Measurement and Statistics: An Examination of Construct Validity. *Psychological Bulletin*, 97, 85-93.

McCall, G. & Simmons, J. L. (eds.). (1969). *Issues in Participant Observation: A Text and Reader*. Reading MA: Addison-Wesley.

Mcdill, E. L., Mcdill, M. S., & Sprehe, J. T. (1972). Evaluation in Practice:

Compensatory Education. In P. H. Rossi and W. Williams(Eds.), *Evaluating Social Programs: Theory, Practice, and Politics*. New York: Seminar Press.

Meidinger, E., & Schnaiberg, A. (1980). Social Impact Assessment as Evaluation Research: Claimants and Claims. *Evaluation Review*, 4, 507-535.

Meltzer, B. N., Petras, J. W., & Reynolds, L. T. (1975). *Symbolic Interactionism: Genesis, Varieties, and Criticism*. London: Routledge & Kegan Paul.

Miller, D. (1983). *Handbook of Research Design and Social Measurement*. New York: Longman.

Mitchell, S. K. (1979). Inter-observer Agreement, Reliability, and Generalizability of Data Collected in Observational Studies. *Psychological Bulletin*, 86, 376-390.

Moustakas, C. (1981). *Rhythms, Rituals and Relationships*. Detroit: Center for Humanistic Studies.

Moustakes, C. (1988).*Phenomenology, Science and Psychotherapy Sydney*, Nova Scotia, Canada: University College of Cape Breton, Family Life Institute.

Applications. Newbury Park, CA: Sage.

Myers, Isabel Briggs & Peter Meyers. (1980).*Gifts Differing*. Palo Alto, CA: Consulting Psychologists.

Neuman, W.L. (2000) *Social Research Methods: Qualitative and Quantitative Approaches*. Boston: Allyn and Bacon.

Neuman, W.L. (1997). *Social Research Methods: Qualitative and Quantitative Approaches*. (*社会研究方法——质化与量化取向.朱柔若译于2000*)台北：扬智.

Oppenhein, A. N. (1966). *Questionnaire Design and Attitude Measurement* New York: Basic Books.

Orne, M. T. (1969). Demand Characteristics and the Concept of

Quasi-controls. In R. Rosenthal and R. L. Rosnow (Eds.), *Artifact in Behavioral Research* (pp.143-179). New York: Academic Press.

Palit, C., & Sharp, H. (1983). Microcomputer Assisted Telephone Interviewing. *Sociological Methods and Research*, 12, 169-189.

Parlett, M. and David H. (1976). Evaluation as Illumination: A New Approach to the Study of Innovatory Programs. In *Evaluation Studies Review Annual*. 1, edited by G. V. Glass. Beverly Hills, CA: Sage.

Patterson, M. L. (1976). An Arousal Model of Interpersonal Intimacy. *Psychological Review*, 85, 235-245.

Patton, M. Q. (1980). *Qualitative Evaluation Methods*. Beverly Hills, Calif: Sage.

Patton, M. Q. (1995). *Qualitative Evaluation and Research Methods*.(质的评鉴与研究.吴芝仪、李奉儒译).台北: 桂冠.

Polanyi, M. 〔1967〕(1983). *The Tacit Dimension*. Magnolia, MA: Peter Smith.(Reprinted)

Posavac, E. J. (Ed.), & Carey, R. G. (1985). *Program Evaluation: Methods and Case Studies*. Englewood Cliffs, N. J.: Prentice-Hall.

Ramos, R. (1981). Participant Observation. In R. M. Grinnell, Jr.(Ed.), *Social Work Research and Evaluation* (pp.348-360). Itasca, IL: Peacock.

Rathje, W. L. (1979). "Trace Measures," in Lee Sechrest (ed.), *Unobtrusive Measurement Today*, Jossey-Bass, San Francisco.

Ray, W. & Ravizza, R. (1985). *Methods Toward a Science of Behavior and Experience*. Belmont, CA: Wadsworth.

Reed, J. G., & Baxter, P.M. (1983). *Library Use: A Handbook for Psychology*. Washington, D. C.: American Psychological Association.

Reichardt, C. S., & Cook, T. D. (1985). *Qualitative and Quantitative Methods in Evaluation Research* (Vol. 1, pp.1-49). Beverly Hills, CA: Sage.

Reynolds, P. D. (1982). *Ethics and Social Science Research*. Englewood Cliffs, N. J.: Prentice-Hall.

Riecken, H. W. & Boruch, R. F. (eds.). (1974). *Social Experimentation: A*

Method of Planning and Evaluating Social Intervention. New York: Academic Press.

Rosenberg, M. J. (1969). The Conditions and Consequence of Evaluation Research. In R. Rosenthal and R. L. Rosnow (Eds), *Artifact in Behavioral Research* (pp.279-349). New York: Academic Press.

Rosenberg, M. (1968). (*调查分析的逻辑*.徐正光、黄顺二合译).台北：国立编译馆.

Rosenthal, R. (1991). *Meta-Analytical Procedures for Social Research*. (*社会研究的后设分析程序*.齐力译于1999).台北：弘智.

Rosenthal, R. & Rubin, D. B. (1982). A Simple, General Purpose Display of Magnitude of Experimental Effect. *Journal of Educational Psychology*, 74, 166-169.

Rosenthal, R. (1976). *Experimenter Effects in Behavioral Research*. New York: Irvington.

Rosenthal, R. (1978). Combining Results of Independent Studies. *Psychological Bulletin*, 85, 185-193.

Rossi, P. H. (1978). Issues in the Evaluation of Human Service Delivery. *Evaluation Quarterly*, 2, 573-599.

Rossi, P. H., Wright, J. D., & Wright, S. R. (1978). The Theory and Practice of Applied Social Research. *Evaluation Quarterly*, 2. 171-191.

Rossi, P., Wright, J. & Anderson, A. (eds.)(1983). *Handbook of Survey Research*. New York: Academic Press.

Royse, D. (1995). *Research Methods in Social Work*. Chicago: Nelson-Hall

Ruckdeschel, R. A. (1985). Qualitative Research as a Perspective. *Social Work Research and Abstracts*, 21, 17-21.

Russell, B. (1948). *Human Knowledge: Its Scope and Limits*. New York: Simon & Schuster.

Rychlak, J. (1977). *Psychology of Rigorous Humanism*. New York: John Wiley & Sons.

Rychlak, J. (1979). *Discovering Free Will and Personal Responsibility*. New

York: Oxford University Press.

Sanders, J. R., & Keithspiegel, P. (1980). Formal and Informal Adjudication of Ethics Complaints About Psychologists. *American Psychologist,* 35, 1096-1105.

Scarr, S., & Weinberg, R. A. (1976). IQ Test Performance of Black Children Adopted by Whole Families. *American Psychologist,* 31, 726-739.

Schaie, K. W. (1965). A General Model for the Study of Developmental Problems. *Psychological Bulletin,* 64, 92-107.

Scheaffer, R., Mendenhall, W. & Ott, L. (1986). *Elementary Survey Sampling* (3rd ed.). Duxbury Press.

Scherer, K., & Ekman, P. (Eds.). (1982). *Handbook of Methods in Nonverba Behavior Research*. Cambridge, Eng.: Cambridge University Press.

Schroeder, L. D., Sjoquist, D. L., & Stephan. P. E. (1986). *Understanding Regression Analysis: An Introductory Guide*. Beverly Hills. Calif.: Sage.

Schwartz, H., & Jacobs, J. (1979). *Qualitative Sociology: A Method to the Madness*. New York: Free Press.

Scriven, M. (1972). *Objectivity and Subjectivity in Educational Research In Philosophical Redirection of Educational Research: The Seventy-First Yearbook of the National Society for the Study of Education,* Edited by L. G. Thomas. Chicago: University of Chicago Press.

Selltiz, C., Wrightman, L. S., & Cook, S. W. (1976). *Research Methods in Social Relations*. New York: Holt, Rinehart & Winston.

Sieber, J. E., & Stanley, B. (1988). Ethical and Professional Dimensions of Socially Sensitive Research. *American Psychologist*, 43, 49-55.

Silverman, I. (1977). *The Human Subject in the Psychological Laboratory*. Elmsford, N. Y.: Pergamon.

Smith, T. W. (1990). The First Straw? A Study of the Origins of Election Polls. *Public Opinion Quarterly*, 54, 1-36.

Stake, R. E. (1981). Case Study Methodology: An Epistemological Advocacy. Pp:31-40 in *Case Study Methodology in Education Evaluation,* edited

by W. W. Welch. Minneapolis: University of Minnesota, Research and Evaluation Center.

Stevens, S. S. (1951). Mathematics, Measurement and Psychophysics, S. Stevens (ed.) *Handbook of Experimental Psychology.* Wiley, New York.

Steward, D.W. (1993). *Secondary Research.* (次级资料研究法.董旭英与黄仪娟合译于2000).台北：弘智.

Stock, M. (1985). *A Practical Guide to Graduate Research.* New York: McGraw-Hill.

Strauss A. & Corbin, J. (1990) *Basics of Qualitative Research: Grounded Theory Procedures and Techniques.* (质性研究概论.徐宗国译).台北：巨流.

Sudman, S. (1976). *Applied Sampling.* New York: Academic Press.

Sudman, S. (1983). Survey Research and Technological Change. *Sociological Methods and Research,* 12, 217-230.

Taylor, J. B. (1977). Toward Alternative Forms of Social Work Research: The Case for Naturalistic Methods. *Journal of Social Welfare*, 4, 119-126.

Thompson, M. S. (1980). *Benefit-Cost Analysis for Program Evaluation.* Beverly Hills, Calif.: Sage.

Thompson, W. C., & Fortess, E. E. (1980). Cost-effectiveness Analysis in Health Program Evaluation. *Evaluation Review*, 4, 549-568.

Tripodi, T., & Epstein, I. (1978). Incorporating Knowledge of Research Methodology into Social Work Practice. *Journal of Social Research*, 2, 11-23.

Turner, J. (1974). *The Structure of Sociological Theory.* Homewood. IL: Dorsey.

Turner, J. H. (ed.)(1989). *Theory Building in Sociology: Assessing Theoretical Accumulation.* Newbury Park, CA: Sage.

Tutty, L., Rothery, M.A. & Grinnell, R. (1996). *Qualitative Research for Social Workers - Phase, Step and Tasks.* MA: Allyn & Bacon.

Uhl, N.P. (1990). Evaluation model and approaches-Delphi technique, In

Walberg, H.J. & Haertal, G.D., *The international encyclopedia of educational evaluation*. Oxford: Pergamon.

Van Maanen, J., Dabbs, J. M., Jr., & Faulkner, R. R. (Eds.). (1982). *Varieties of Qualitative Research*, Beverly Hills, CA: Sage.

Volk, K. (1993). Curriculum development using the Delphi technique. *The Technology Teacher,* January, 35-36.

Wallance, W. (1971). *The Logic of Science in Sociology.* Chicago: Aldine-Atherton.

Watts, T. D. (1981). Ethno-methodology. In R. M. Grinnell Jr. (Ed.), *Social Work Research and Evaluation* (pp.361-372). Itasca, IL: Peacock.

Watts, T. D. (1985). Ethno-methodology. In R. M. Grinnell Jr. (Ed.), *Social Work Research and Evaluation* (2nd ed. ，pp.357-369). Itasca, IL: Peacock.

Webb, E., Campbell, S., Schwartz，R., & Sechrest, L. (1966). *Unobtrusive measures: Non-reactive Research in Social Science*. Chicago: Rand McNally.

Webb, T. M. (1996) . *The perception of African American community leaders regarding curriculum planning for minority students: A Delphi study*. UMI Dissertation Services, D'Youville College. Measures: Non-reactive Research in Social Science. Chicago: Rand McNally.

Weinbach , R. W. , & Grinnell , R. M. , Jr. (1987). *Statistics for Social Workers*. White Plains , NY: Longman.

Weiss, C. (1972). *Evaluation Research*. Englewood Cliffs, NJ: Prentice-Hall.

Westinghouse Learning Corporation & Ohio State University. (1969). *The Impact of Head Start: An Evaluation of Effects of Head Start on Children's Cognitive and Affected Development*. Vols. 1 and II. (Order No. PB 184329). Springfield, Va.: Clearinghouse for Federal Scientific and Technical Information, U. S. Department of Commerce.

Williams, B. (1978). *A Sampler on Sampling*. New York: John Wiley.

Williams, G. A. (1989). Enticing Viewers: Sex and Violence in TV Guide

Program Advertisements. *Journalism Quarterly*, 66 , 970-973.

Wilson , E . B. J. (1952). *An Introduction to Scientific Research*. New York: McGraw-Hill.

Yamaguchi K. (1991). *Event History Analysis*.（事件史分析.杜素豪与黄俊龙合译于2001).台北：弘智.

Yegidis B. L. & Weinbach, R. W. (1996). *Research Methods for Social Workers*. Boston: Allyn and Bacon.

Yin, R.K. (1994) . *Case Study Research*.（个案研究.尚荣安译于2001).台北：弘智.

Zaltman, G. (1979). Knowledge Utilization as Planned Social Change. *Knowledge: Creation , Diffusion , Utilization*, 1, 82-105.